OPTIMIZATION OF PRIMARY AND ENHANCED SHALE OIL RECOVERY USING STATISTICAL DESIGN OF EXPERIMENTS AND DATA ANALYTICS

A Dissertation

by

MINA ROMANY SHAKER MOUSA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hisham Nasr-El-Din
Co-Chair of Committee,	Stephen A. Holditch
Committee Members,	Daren B. H. Cline
	Mahmoud El-Halwagi
Head of Department,	Jeff Spath

December 2018

Major Subject: Petroleum Engineering

ABSTRACT

The application of statistical analysis and data analytics is not widespread in the petroleum industry, but is gaining recognition and will be applied more in the future. Applying these tools could improve the accuracy and precision of results, which could have a positive effect on the decision-making process. Therefore, both statistical analysis and data analytics constituted the core of this research to generate an efficient systematic workflow that was applied to study one of the major problems that characterize the complex shale reservoirs, namely how to improve their low recovery factor (RF).

A simulation-based design-of-experiments (DoE) workflow was developed to optimize the design of four recovery schemes to maximize the low RF in Eagle Ford shale. These schemes are primary production (PP), waterflooding, continuous miscible gas flooding, and miscible gas huff 'n' puff. The workflow was used to pinpoint the optimum spots in the multidimensional variable space. Using an innovative injection pattern that relies on alternating injection and producing fractures along the same lateral, continuous miscible gas flooding was found to have the highest potential to maximize RF. Developing this injection pattern might be the next breakthrough to boost the low shale RF.

Adequate representation of the complex and challenging shale reservoirs using numerical simulation necessitates making many uncertain assumptions, which could affect the reliability of its results. Leveraging less presumptive techniques like data analytics (DA) is required to validate modeling results. Both DA and DoE were applied on a Bakken shale PP case study. Results showed that RF has a physical limit that cannot be exceeded

by merely optimizing PP, which can only accelerate oil production and cash flow. Metamodeling was used for optimization and quantifying the effects of uncertainties in reservoir characteristics and design variables.

The main purpose of this research is to illustrate the use of DoE in the oil and gas industry. DoE provides a systematic research framework that can be leveraged for an efficient exploration of the multidimensional variable space to pinpoint the optimum spots. This framework produces statistically-based conclusions and data-driven facts, which could improve the objectivity of the decision-making process.

We hope that this work could help encourage petroleum researchers and engineers to incorporate statistics at the heart of industrial and academic research and problem-solving tools, and to learn and apply DoE regardless of the type of experiments that they conduct; physical experiments, simulation runs, or field trials.

DEDICATION

To my daughter Emily, for her smile.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Nasr-El-Din, for being a great mentor during my time at Texas A&M, for believing in me, and for his understanding and flexibility.

I would also like to thank my committee co-chair, Dr. Holditch, for the endless knowledge and support he provided me, especially during the hard times.

I would also like to acknowledge my committee members, Dr. Cline and Dr. El-Halwagi, for their guidance and time.

I am truly grateful that to Texas A&M University for the great learning environment it provides to its students. Thanks to my friends, colleagues, faculty, and staff for making my time at Texas A&M University a great experience. Thanks also to Dr. Maggard for providing me an opportunity to work with him as a teaching assistant and an instructor.

Finally, thanks go to my mother, father, and brothers for their encouragement, and to my wife and daughter for their patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Nasr-El-Din (advisor) and Holditch (co-advisor) of the Department of Petroleum Engineering, Professor Cline of the Department of Statistics, and Professor El-Halwagi of the Department of Chemical Engineering. All work conducted for the dissertation was completed independently by the student. Data used in Chapter 5 was obtained from DrillingInfo and Wood McKenzie.

Funding Sources

Graduate study was completed without outside financial support.

NOMENCLATURE

A-LASSO	Adaptive Least Absolute Shrinkage and Selection Operator
A-EN	Adaptive Elastic Net
AICc	corrected Akaike Information Criterion
ANN	Artificial Neural Networks
ARE	Absolute Relative Error
Bbl	Barrel
BF	Bootstrap (or Random) Forests
BHP	Bottom-Hole Pressure
BIC	Bayesian Information Criterion
BT	Boosted Trees
CAPEX	Capital Expenses, USD MM
CON	Miscible Continuous Flooding
Cum365	Cumulative 90-day production (bbl)
Cum90	Cumulative 365-day production (bbl)
DA	Data Analytics
DoE	Design of Experiments
DP	Dykstra Parsons coefficient, fraction
EN	Elastic Net
ENS	Ensembles
EOR	Enhanced Oil Recovery
EUR	Estimated Ultimate Recovery (million barrels or mmbbl)
FDR	False Discovery Rate
HNP	Huff 'n' Puff

IOR	Improved Oil Recovery
IP	Initial Production for the month with the maximum daily production rate (barrels/day or b/d)
KNN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Linear Regression
MAE	Mean Absolute Errors
MC	Monte Carlo
MdAE	Median Absolute Error
MdARE	Median Absolute Relative Error
MAR	Missing At Random
MARE	Mean Absolute Relative Errors
MCAR	Missing Completely At Random
MI	Multiple Imputation
MMP	Minimum Miscibility Pressure, psi
MNAR	Missing Not At Random
Mscf	Thousand standard cubic feet
NPV	Net Present Value
OFAT	One Factor at A Time
OPEX	Operating expenses, USD MM
PP	Primary Production
PRESS	Predicted Sum of Squares
R&D	Research and Development
RE	Relative Error
RF	Recovery Factor

RMSE	Root Mean Square Error
RR	Ridge Regression
RSE	Relative Squared Error
RSS	Residual Sum of Squares
RT	Regression Trees
SL	Supervised Learning
Tbbl	Trillion barrels
TF	Transverse Fractures
Trn	Training Set
Tst	Test Set
UCL	Upper Control Limit
USD MM	Millions of US dollars
VIF	Variance Inflation Factor
Vld	Validation Set
WF	Waterflooding

c_i	Coefficient (best fit) or weight (weighted average) of i^{th} technique
Cp	Mallow's Cp statistic
H_0	Null hypothesis
H_1	Alternative research hypothesis
k	Matrix permeability, md
k_{eff}	Effective fracture block permeability, md
k_f	Fracture permeability, md
n	Number of data points in the training or validation set
R^2	Coefficient of determination

R^2_{adj}	adjusted coefficient of determination
w	Degree of miscibility
w_{eff}	Effective fracture width, ft
w_f	Actual fracture width, ft
$\bar{y}$	Average of n actual responses
$\hat{y}_{ENS}$	Estimated ensemble value for any of the four responses
$\hat{y}_i$	i^{th} estimated response
y_i	i^{th} actual response
α	Significance level
β_i	Regression coefficient
ε	Error
μ	Mean of normal distribution of ln (k)
σ	Standard deviation of normal distribution of ln (k)
φ_{eff}	Effective fracture porosity, fraction
φ_f	Fracture porosity, fraction
φ_m	Matrix porosity, fraction

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION AND LITERATURE REVIEW

The application of statistical analysis and data analytics is not widespread in the petroleum industry, but is gaining recognition and will be applied more in the future. Applying these tools could improve the accuracy and precision of results, which could have a positive effect on the decision-making process. Therefore, both statistical analysis and data analytics constitute the core of this research to generate an efficient systematic workflow that can be applied to study one of the major problems that characterize the complex shale reservoirs, namely how to improve their low recovery factor (RF).

The Energy Information Administration (EIA 2013) estimated that the total risked shale oil in place is around one trillion barrels (Tbbl) for the US, and around seven Tbbl for the world. These numbers might be highly conservative due to lack of data (Maugeri 2012). Over the past decade, the technological improvements in horizontal wells and multistage hydraulic fracturing have opened access to these vast oil resources, which were earlier considered uneconomical. In 2015, the US produced about 4.9 million barrels per day (bpd) of crude oil from shale oil resources, which represented around 52% of its total crude oil production (EIA 2016a). However, the RF from shale is low, ranging from 1 to 10% of the original oil in place (Clark 2009; Hoffman 2012; Chen 2013; Gamadi 2014; Kathel and Mohanty 2013; Joshi 2014; Morsy 2014; Sheng and Chen 2014; Sheng 2015). In the last decade, several hundred studies (experimental and numerical simulation) investigated several improved oil recovery (IOR) techniques to improve the low RF of shale plays. Alfarge et al. (2017) reviewed many of such studies. The studied IOR

techniques include waterflooding (WF), huff and puff miscible gas injection (HNP), and miscible continuous gas flooding (CON).

1.1 Waterflooding

Many researchers applied simulation to evaluate waterflooding (WF) potential to improve oil recovery in shale reservoirs (**Table 1.1**). Some found out that WF could have large impact on RF (Wang et al. 2010; Morsy et al. 2013; Sheng and Chen 2014; Kong et al. 2016). On the other hand, there are some other researchers who found out that WF does not produce appreciable improvement in RF (Iwere et al. 2012; Dong and Hoffman 2013; Kurtoglu 2014; Nguyen et al. 2015). The ultra-low permeability of shale is probably the main cause of the very poor injectivity during WF (Chen 2013; Fai-Yengo et al. 2014; Sheng and Chen 2014; Nguyen et al. 2015). Although the general trend in the literature is to rule out WF as a potential IOR technique (Sheng 2015), some contradiction still exists in the reported results. In addition, the number of WF studies is small. Thus WF is considered in this work just for the purpose of comparison.

Reference	Primary RF, %	WF RF, %	RF increment, %
Kong et al. (2016)	-	-	up to 10
Morsy et al. (2013)	12	up to 21	9
Sheng and Chen (2014)	6.5	11.9	5.4
Wang et al. (2010)	3.64	8.3	4.66
Kurtoglu (2014)	-	-	1.12
Dong and Hoffman (2013)	-	-	0.94
Nguyen et al. (2015)	-	-	0.45
Iwere et al. (2012)	-	-	0.3

Table 1.1—RF improvement in the literature for WF.

1.2 Miscible Gas Enhanced Oil Recovery

Based on the Enhanced Oil Recovery (EOR) screening guidelines developed by Taber (1983) and Taber et al. (1997), Iwere et al. (2012) concluded that gas injection methods are the most suitable EOR technique for Bakken and Three Forks shale formations. According to the available literature on EOR in shale oil plays, Sheng (2015) concluded that miscible gas injection is more efficient than WF. **Table 1.2** summarizes some of the results in the literature.

Six of the eight cases (Wang et al. 2010; Chen 2013; Kurtoglu 2014; Sheng and Chen 2014; Fragoso et al. 2015; Nguyen et al. 2015; Sheng et al. 2015; Fragoso et al. 2017) that investigated both CON and HNP agree that the former gives higher recovery. However, the general trend in the literature considers HNP to have a higher potential to improve RF (Sheng 2015). This might be attributed to the higher risk of early gas breakthrough through the natural fractures system in case of CON, and to the ultralow permeability, which could make CON more difficult to conduct (Sheng and Chen 2014; Sheng 2015). Thus my research considers both techniques. Innovative injection patterns that apply both injection and production from the same well will be investigated.

1.3 Scope and Organization

Most of the published work about IOR in shale used one factor at a time (OFAT) to conduct research. However, Design of Experiments (DoE) is a more accurate and precise method. It leads to a better understanding of the system performance over a wide range of conditions, enhances response prediction in the variable space, and produces more valid conclusions (Daniel 1973; Czitrom 1999; Santner et al. 2003; Box et al. 2005; Mathews

Reference	Permeability	Primary RF, %	Gas inj. RF, %	RF increase, %	Inj. scheme
Wang et al. (2010)	0.04 and 2.5 md	3.64	up to 36.2%	up to 32.57	CON
			28.9	25.26	HNP
Iwere et al. (2012)	13 µd to 1.06 nd	6.4	7	0.6	CON
Chen (2013); Sheng and Chen (2014)	100 nd and 0.001 md	6.46	up to 73.65 (≈ 100% for 0.001 md)	up to 67.19	CON
			up to 14.42	up to 7.96	HNP
Dong and Hoffman (2013)	40 µd	5.42	up to 30.01	up to 24.59	CON
Wan et al. (2013a)	100 nd	6.5	up to 60	up to 53.5	HNP
Wan et al. (2013b)	100 nd	5.5	up to 70	up to 64.5	HNP
Chen et al. (2014)	0.01 md	5.83	5.59	-0.24	HNP
Fai-Yengo et al. (2014)	80 µd			up to 28.5%	CON
Kurtoglu (2014)	0.031 md	5.24	5.81 (max.)	up to 0.57	CON
			5.43 (max.)	up to 0.19	HNP
Yu et al. (2014)	5 µd	≈ 14	≈ 23.4 (max.)	up to 9.4	HNP
Li et al. (2015)	500 nd			up to 70.07	HNP
Fragoso et al. (2015)	250 nd	10.62	up to 40.63	up to 30.01	CON
Fragoso et al. (2015)	250 nd	10.62	up to 32.55	up to 21.93	HNP
Nguyen et al. (2015)	500 nd	11.58	15.08 (max.)	up to 3.50	CON
			13.45 (max.)	up to 1.87	HNP
Sheng et al. (2015)	220 nd	3.2	4.3	1.1	CON
			7.1	3.9	HNP
Vinassa et al. (2015)	250 nd	10	53	43	HNP
Li et al. (2016)	500 nd	5.656	up to 31.52	26.864	HNP
Chen and Gu (2017)	10 µd	7.47	up to 26.65	up to 19.18	HNP
Fragoso et al. (2017)	100 nd	6.9	up to ≈ 27.5	up to 20.6	CON
			up to ≈ 35.9	up to 29.0	HNP
Kanfar et al. (2017)	5 µd	8	up to 11.4	up to 3.4	HNP
Li et al. (2017)	300 nd	6.08	up to 18.3	up to 12.22	HNP
Zhu et al. (2017)	10 µd	17.1	32.8	15.7	CON
	1 µd	9.4	21.9	12.5	CON

Table 1.2—RF improvement in the literature for miscible gas injection.

2005; Eriksson 2008; Wu and Hamada 2009; Montgomery 2012). A few studies handled shale oil production using DoE or its basic principles (Ghaderi et al. 2012; Kalra and Wu 2015; Popov et al. 2016; Chen and Gu 2017; Joslin et al. 2017). Nevertheless, they dealt with it from an EOR perspective. To the best of our knowledge, no one has studied primary production (PP) or WF in shale reservoirs using DoE (detailed discussion about

DoE is provided in Chapter 2). Thus one of the main goals of this work is to develop a DoE-based simulation workflow to determine the potential of optimizing the design of four recovery schemes (PP, WF, CON, and HNP) to maximize the low RF in Eagle Ford shale (Chapter 3). This is done by achieving the following objectives:

1. Estimate the effects of design variables on RF more accurately and more precisely
2. Define the most important variables that affect the RF of PP, WF, CON, and HNP
3. Specify if PP optimization is enough to improve the low primary RF
4. Compare the four recovery schemes to identify the most prospective one
5. Determine which miscible injection scheme (continuous flooding vs. cyclic injection) is better
6. Specify the injection gas that yields the best economic performance
7. Determine the optimum combination of important design variables that maximizes the RF for selected recovery schemes
8. Evaluate the potential of an innovative injection scheme that uses alternating injecting and producing fractures along the same lateral

Since this workflow is based on simulation, adequate representation of the complex and challenging shale reservoirs necessitates making numerous uncertain assumptions. In addition, reservoir simulation can be time consuming. Thus other less presumptive methods are required to validate simulation results (Mohaghegh et al. 2017). Data Analytics (DA) offers a powerful tool in this respect because it avoids making the physical assumptions associated with simulation. Therefore, using DA could reduce the

uncertainties associated with shale oil production, which could reduce risk and improve the robustness to the decision-making process.

For comparison purposes, both DA and simulation/DoE workflows are applied. The DoE-based simulation workflow is applied to PP in Bakken shale (Chapter 4). It considers both controllable design variables as well as uncontrollable state variables (e.g., reservoir permeability). In addition to the first three objectives previously stated for Chapter 3, this chapter adds the following objectives:

1. Compare the relative importance of two categories of variables: controllable and uncontrollable variables
2. Understand how the dominance of the effects of the most important variables on production economics change over time
3. Assess the effects of uncertainties in different design and state variables on oil production economics

Then, Chapter 5 harnesses the potentials of DA and applies it to a big dataset for Bakken shale to produce statistically-based conclusions and data-driven facts. The objectives here are as follows:

1. Identify key factors that distinguish good wells from poor-performing ones
2. Mine data-driven insights from the available big data and translate them into better understanding of shale reservoirs and improving their production performance
3. Compare DA conclusions to those based on reservoir simulation and DoE (Chapter 4) to evaluate the validity of the results from the latter and assess the performance of simulation

2. DESIGN OF EXPERIMENTS IN PETROLEUM ENGINEERING: HISTORY AND CURRENT CHALLENGES

2.1 Introduction

Engineering research is a systematic approach to understand, design, and/or optimize a system (product or process) through collecting, analyzing, and interpreting information obtained by experimentation (National Academy of Engineering 1995; Kothari 2004; Leedy and Ormrod 2013). A widely applied type of research is analytical research, which is usually concerned with cause-effect relationships. This type of research is performed to evaluate the influence of the explanatory variable(s) on the response variable(s) by conducting experiments (Montgomery 2012). Experiments here might be laboratory tests, modeling runs, or field trials.

Due to the inherent variability in experimentation, engineers and scientists need to use statistics. In addition, with the increasing cost of experimentation, research should be done as efficiently as possible by maximizing the amount of useful information per experiment. Design of Experiments (DoE) is used to maximize research efficiency. However, in general, as well as in the petroleum industry, engineers and scientists extensively apply two other approaches for evaluating cause-effect relationships; best-guess and one-factor-at-a-time (OFAT) (Montgomery 2012).

2.1.1 Best-Guess Approach

Best-guess starts with selecting an arbitrary combination of levels of factors. Experiments are then conducted to see what happens. The results and observations of the

preliminary experiments will direct the experimenter to the next combination of levels of factors for the next experiment or set of experiments. It often works well when the experimenters have good knowledge and experience with the system. One major disadvantage of this approach however is that, in searching for the desired outcomes, it could continue almost indefinitely, changing the levels of one, two or several factors for the next test, based on the results of the current test. In addition, there is no guarantee for success. Another drawback is that the researcher might decide to stop testing if any of the best-guess attempts produces acceptable results. This does not mean, however, that the best solution has been found (Montgomery 2012). Moreover, it is often difficult to describe how the optimal settings of factors are obtained, quantify factor effect, or develop an empirical prediction model (Lye 2002).

2.1.2 One-Factor-at-A-Time (OFAT)

The first step in OFAT (sometimes called OVAT, one-variable-at-a-time, or COST, changing one single variable at a time) is to select a base case, i.e., a certain set of levels for each factor. Then, one factor is changed over its range while holding all other factors constant at their base-case levels. The process repeats for all factors. Eventually, the experimenter constructs a series of graphs showing how each factor affects the response variable with all other factors held constant. However, the main effects estimated are conditional effects (not general effects averaged over the entire design space). The main disadvantage of OFAT is that it does not account for interactions, which is very common in many engineering experiments, leading to poor and sometimes even wrong conclusions. Many experimenters (including the experienced ones) are not familiar with this issue, or if

they are, they do not know how evaluate it. Therefore, “there are a large number of engineering phenomena resulting from interactions and other complexities that could be revealed by experimental design and are waiting to be discovered and exploited,” (Box 2001). Another drawback appears while searching for the optimum, where OFAT is inefficient in the sense that there is a very low probability that the experimenter can hit the optimum conditions. This is true even when the number of factors is as low as two. It becomes increasingly unreliable and inefficient as the number of factors increases. It can reach the optimum only when there are no interactions or interactions are very small compared to the smallest of the main effects. Like the best-guess approach, OFAT could mistakenly lead the experimenter into thinking that the optimum is reached. Unfortunately, a lot of engineers and scientists still think that OFAT and best-guess approaches are the standard research tools to explore cause-effect relationships (Koselka 1996; Lye 2002; Box et al. 2005; Wu and Hamada 2009; Montgomery 2012).

Despite its many drawbacks, OFAT can still be used in some very specific circumstances. It could be economic to apply when the effects are more than four times the standard deviation of the random error. However, the effect estimates could be biased if large interactions exist. In addition, OFAT does not consider randomization and uncontrollable time trends will have a detrimental effect on the results, especially when experiments take a long enough time for such trends to be appreciable (Daniel 1973). Thus OFAT can be applied only when the errors are small and the experiments are quick. In deterministic computer modeling however, time trends and random errors do not exist, which could make OFAT a suitable approach when the number of modeling scenarios is

limited to approximately the number of factors (Frey et al. 2003). Frey and Wang (2006) showed that adaptive OFAT can be better than resolution III fractional factorial designs that have the same number of experiments (adaptive OFAT requires n(k-1)+1 experiments with k levels for each of n factors), especially when interactions are large (quarter of the main effects) or errors are small. They also demonstrated that adaptive OFAT can provide substantial improvements in early stages, which could be an advantage if the experiments need to be terminated before the plan is completed.

Experience and knowledge about the system are very important to decide how the factors will be set to provide a good point for starting design space exploration, and how the factors will be changed thereafter by OFAT (Wu and Hamada 2009). These are not always available however, especially for new systems. Even if they are available, applying a more systematic approach with an intelligent sampling of the variable space can provide more efficient and effective methodology to achieve study objectives by delivering a larger amount of more accurate and more precise information per experiment. This more systematic approach is design of experiments, DoE.

2.1.3 DoE vs. OFAT

Statistical design of experiments, or DoE, is almost always better than the inefficient best-guess and OFAT approaches for examining cause-effect relationships (Fisher et al. 1990; Montgomery 2012; Anderson and Whitcomb 2017). **Fig. 2.1** compares both approaches in a simple optimization problem involving two factors. OFAT is a non-systematic trial-and-error approach, which depends on luck, intuition, and experience (Antony et al. 2003). This is apparent in Fig. 2.1a. In addition, OFAT can mislead the

experimenter to believe that point (10) is the optimum. On the other hand, changing all factors simultaneously based on DoE provides more decisive information regarding the location of the optimum (Fig. 2.1a) with less experiments - five experiments (four full factorial plus one center-point) for DoE compared to 12 for OFAT. Thus DoE provides more higher-quality information with less resources.

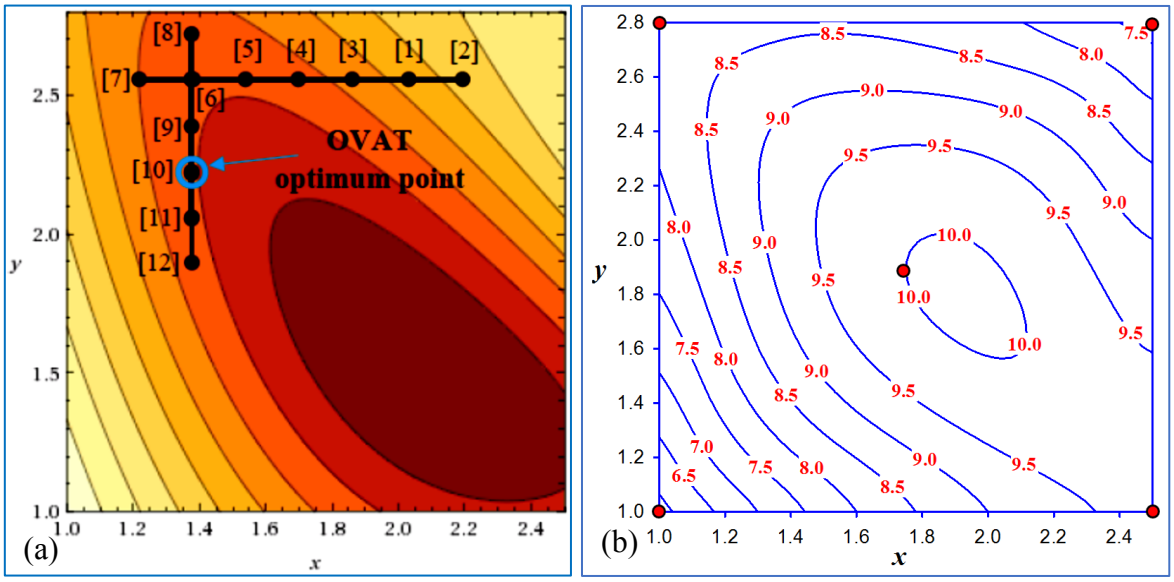


Fig. 2.1—Comparison between OFAT and DoE: (a) OFAT (11 points); (b) DoE (5 points for full factorial with centerpoint). The darkest region in (a) represents the optimum operating conditions.

Fig. 2.1 demonstrates a simple problem where only two factors are involved. How would OFAT work when we have more than that? Exploring the variable space will suffer. DoE can handle such cases more efficiently. Below are some of the reasons why engineers and scientists should use DoE (Daniel 1973; Czitrom 1999; Box et al. 2005; Mathews 2005; Eriksson 2008; Wu and Hamada 2009; Hibbert 2012; Montgomery 2012):

a. DoE handles more efficiently situations where the objective function depends on many explanatory variables (factors).
b. DoE incorporates randomization, which is ruled out in OFAT making its estimates biased because of time trends.
c. DoE consumes less resources for obtaining more higher-quality information.
d. DoE defines the number of experiments required before actual experimentation.
e. DoE does not require prior knowledge of the physics that governs the system.
f. DoE evaluates the effect of one factor at more than one level of other factors. Thus it provides more accurate and more precise factor effects by reducing the bias and variability in their estimates. This leads to a better understanding of the system performance over a wide range of conditions, enhances response prediction in the factor space, and produces valid conclusions.
g. DoE systematically estimates the interaction between factors. OFAT cannot do that.
h. DoE provides a more efficient optimization process since the whole multidimensional variable space is searched for the optimum.
i. DoE leads to conclusions that are based on facts, which improves the objectivity of the decision-making process.

In spite of such facts, the vast majority of petroleum engineers, especially those working in the field and those focusing on lab experiments, still use the inefficient OFAT and best guess approaches. Some situations could be justified as stated earlier. However, this is not the case for most analytical research studies. OFAT is still widely used because

it is intuitive. However, intuition is not always correct or enough, especially when it does not have strong foundations. Another reason is that many researchers have little or no knowledge about DoE, or have little or no enthusiasm to learn and apply DoE (Daniel 1973). Furthermore, investigators prefer OFAT because they can see and react to the experimental results faster, which - according to cognitive psychology - can help them more readily understand the physics of the system under investigation (Daniel 1973; Wu and Hamada 2009). Also using DoE could seem intimidating in the beginning when one determines the required number of experiments. This is not true for OFAT. However, the systematization, the efficiency of obtaining more higher-quality information, and the higher potential of solving various problems outweigh the efforts (Weissman and Anderson 2015).

2.2 Design of Experiments (DoE)

The National Institute of Standards and Technology (NIST), and the Semiconductor Manufacturing Technology (SEMATECH) consortium (NIST/SEMATECH 2016) define DoE as "a systematic, rigorous approach to engineering problem-solving that applies principles and techniques at the data collection stage so as to ensure the generation of valid, defensible, and supportable engineering conclusions. In addition, all of this is carried out under the constraint of a minimal expenditure of engineering runs, time, and money." DoE is the "gold standard" approach to establish cause-effect relationships (Santner et al. 2003). Statistical design and analysis of experiments should be placed at the heart of the research process.

2.2.1 DoE Applications

Here is a brief description of DoE applications. More details can be found elsewhere (Eriksson 2008; Wu and Hamada 2009; Montgomery 2012; NIST/SEMATECH 2016). Neither best-guess nor OFAT is efficient for any of these applications.

1. Characterization: It is faced when the researcher wants to get more familiar with a new system that is not well-understood. No more than 10% of the available resources is usually spent in this stage.
2. Treatment comparison: The focus here is on a single factor to determine how its levels (treatments) affect the response. This might be similar to OFAT in some sense. However, the conclusions are more objective because it is based on sound planning and execution of experiments, and supportable inferential statistical analysis.
3. Screening: It is conducted to identify the most significant variables - along with their ranges - from a pool of several variables, some of which might not be significant for the response. Once identified, these significant variables can then be used in the more thorough extensive applications that follow.
4. Optimization: Clearly, the objective is to optimize (minimize or maximize) the response(s) using optimization methodologies. An intelligent approach is to apply a sequential strategy which gets factor settings closer and closer to the optimum region. Once reached, a more thorough exploration using DoE (sometimes with metamodeling) can be applied to locate the optimum. In case of several responses, there is usually a conflict between the various objective functions so that they all

cannot be fulfilled simultaneously. Here the analyst should prioritize the responses based on their importance. The desirability function is a common technique to optimize multiple response processes (Harrington 1965).

A special case of optimization applied in reservoir engineering is history matching, where the objective function to be minimized is some measure of the error (difference between actual history and model output).

5. Robustness: DoE is done here to evaluate the sensitivity of a system (product or process) to small changes that might happen to factor settings and to noise. If the system is not robust, then the variable settings should be changed to reach a robust region in the design space.
6. Uncertainty analysis: The objective is to evaluate the effects of uncertainty in the variables on the response so that decision-making process could consider the probabilistic random nature of the problem under investigation.
7. Metamodeling: It is a combination of mathematical and statistical tools used to build a proxy with a high response-prediction power. Other names for proxy that could be found in the literature include auxiliary model, surrogate model, response surface (RS), and emulator (Jin et al. 2001; Giunta et al. 2003; Kleijnen et al. 2005; Yeten et al. 2005). Conventionally (in physical experiments), response surface methodology (RSM) is a more common term than metamodeling. However, metamodeling is more widely used in computer experiments. In addition to RSM, which employs regression-based, low-order polynomials (Box et al. 2005; Myers et al. 2009), metamodeling includes various other techniques that could be more

appropriate than RSM for computer experiments. Examples include artificial neural networks (Smith 1993; Cheng and Titterington 1994), multivariate adaptive regression splines (Friedman 1991), radial basis function (Hardy 1971; Dyn et al. 1986), inductive learning (Evans and Fisher 1994), and kriging (Sacks et al. 1989b; Booker et al. 1999).

Metamodeling is very useful, especially when the experiment is expensive. It has become a very important tool in general engineering design (Goodwin and Powell 2012). Metamodeling can be used in conjunction with DoE for the following purposes (Simpson et al. 2001c; Wang and Shan 2006; Eriksson 2008; Wu and Hamada 2009; Montgomery 2012; Law 2014; Ludvigsen and Le 2015):

1. Sensitivity analysis: It provides more insight into the effects of various variables on the response.
2. Design space exploration and response prediction: It helps explore the behavior of the response throughout the multi-dimensional variable space, especially for variable settings that were not tested.
3. Optimization: It helps determine the combination of various variables that optimize the response of interest.
4. Uncertainty analysis: It helps evaluate the impact of uncertain variables on the response through Monte Carlo simulation.
5. Saving time and effort: For the previous four applications, many runs (could be millions in case of computer experiments) might be required. This could be

impractical, if not impossible. A proxy can be used in such cases because it can evaluate the response in a small fraction of a second.

DoE is not restricted to physical experiments. In fact, running computer experiments occasionally provides some advantages over physical experiments. Sometimes it is practically impossible, time consuming, and/or too expensive to run physical experiments and therefore, computer experiments could be a feasible alternative. Computer experiments also can handle a lot more input variables because they are easier to change. Furthermore, they have the ability to account for considerable higher-order interactions, which are usually important in modeling, without the need to make assumptions about RS nature which is usually more complex (Simpson et al. 1997; Santner et al. 2003; Sanchez 2005). At least, modeling will provide a good start to increase the pace of optimization with less physical experiments to be conducted on a system. Thus industries are shifting their experimentation to the computer rather than physical experiments. Instead of utilizing the advancements in creating faster computer models, they are more used to drive the models to be more complicated to better match the physical processes. This makes the modeling process tedious and time consuming. Thus DoE is required to make system exploration and/or optimization faster and more efficient (Joseph 2016). Although DoE is not widely or effectively used in modeling in many disciplines, (Kleijnen et al. 2005), it is mainly applied in modeling (not in physical experiments) in petroleum engineering. As we shall see, 76.35% of onepetro publications that implemented DoE are modeling studies.

2.2.2 DoE Guidelines

One comprehensive study would necessitate a strong background knowledge about the system under investigation. However, this might lead to frustration if the assumptions made based on such knowledge are not true. Using a sequential approach of experimentation, in which all objectives are sequentially fulfilled in a series of smaller experimental plans, would be more useful and save resources. Below are some general steps for implementing DoE to solve cause-effect problems. More details can be found elsewhere (Coleman and Montgomery 1993; Wu and Hamada 2009; Montgomery 2012).

1. Define the problem statement and the goal of the study. The research problem should be clearly delineated. Its goal can be divided in several objectives (mostly one of the seven previously discussed applications). They should be clear and achievable based on the available resources.
2. Plan for experiments properly. Failure to do so leads to wasting resources and poor results, and complicates the analysis (if it can still be done). Planning can be done as follows:
    i. Select the response(s). They should be measurable, preferably continuous. First, the response's measurement device should be calibrated to ensure that the measurements are accurate (accuracy is a measure of bias, which is the difference between true and measured values). After calibration, a gauge repeatability and reproducibility (R&R) study can be done to ensure that the measurement device is precise and reliable. Repeatability and reproducibility are measures of variation caused by the measurement device and the operator,

respectively. When repeatability and reproducibility are assured, the system is said to be precise and reliable, respectively.

ii. Define factors and ranges. All potential factors that might affect the process should be considered. A cause-effect (fishbone) diagram can help organize the possible factors. The ranges should be determined based on available knowledge about the system and the technical feasibility of achieving such ranges. The number of levels should be carefully chosen. This would depend on the DoE objective (screening, optimization, etc.) and the nature of the effects (linear vs. nonlinear).

iii. Consider replication, blocking, and randomization (**Table 2.1**). These three principles are applied to enhance the signal and reduce the noise to improve test power. Randomization is used to control bias, which is the systematic influence of unknown extraneous (uncontrollable and unmeasurable) nuisance variables. Replication is used to estimate experimental error and control noise, which is unsystematic influence of the unknown extraneous nuisance variables. If these uncontrollable factors are measurable, analysis of covariance can handled them. Blocking is used to control the influence of recognized controllable nuisance variables that are not of interest. Balanced designs should also be promoted whenever resources allow. Balanced designs facilitate the analysis of the results, maximizes the power of the test, and gives more tolerance to model analysis and analysis of variance (ANOVA) against equal variance assumption.

3. Choose a proper screening DoE to determine which of the multitude of possible factors significantly affects the response.

	Nuisance factor			
Degree of control	Controllable	Uncontrollable		
Measurability		Measurable	Unmeasurable	
Influence			Systematic (Bias)	Unsystematic (noise)
How treated	Blocking	Analysis of Covariance	Randomization	Replication

Table 2.1—Categories of nuisance factors and how to treat their influence.

4. Conduct the experiments. Trial runs might be required in the beginning to get familiar with the system and make sure that there are no problems. In addition, the results of the preliminary runs should make sense. This is especially true for modeling where the model should be well-tuned to avoid "garbage in, garbage out". Prior planning to this stage of running experiments is critical for success. Otherwise, the results could be frustrating.
5. Analyze the experimental results to determine the significant variables. The statistical analysis would be easy if the experiments are properly designed and conducted. The experimenters might change the ranges and/or add more factors as appropriate based on what they learn in the screening stage. Objectives might be added, canceled, and/or modified.
6. Plan for another DoE if needed. An optimization DoE might be chosen to focus on the significant variables and determine which levels they should be set at to reach the optimum response(s). A sequential DoE approach might be useful in this case.

7. Fit a metamodel to the data if required.
8. Implement a robustness DoE if needed to make sure that the process is robust against the extraneous variation which cannot be controlled, and against small changes in factor settings.
9. Analyze the results and make sure that the objectives are fulfilled. Confirmation experiments can be conducted to support the chosen factor settings.

2.3 DoE in Petroleum Engineering

In the petroleum industry, many papers provide some details about various types of DoE and metamodeling techniques (Saxena and Vjekoslav 1971; Peng and Gupta 2004; Yeten et al. 2005; Amudo et al. 2009; Haight 2010; Wolff 2010; Goodwin 2015; Shams 2016). Detailed discussion of DoE and metamodeling is outside the scope of this chapter. Interested readers are referred to the previously cited papers, as well as other references cited in this chapter (for example, Montgomery (2012) for DoE in physical experiments, and Santner et al. (2003) for DoE in computer experiments).

DoE is generally overlooked in petroleum engineering publications and is sometimes applied incorrectly. By searching for the keywords “experimental design” and “design of experiment” in SPE online library (onepetro.org), Chidi et al. (2014) found that applying DoE in the petroleum industry has been steadily increasing, especially after 1990. However, most of these papers did not actually apply DoE, and there are no much details regarding how the rest applied DoE or the nature of such studies. Thus it is one of the main goals of this chapter to address such shortages in the literature to see where the petroleum industry stands.

An extensive literature review is conducted through onepetro for all the papers that have any of the keywords “design of experiments”, “design of experiment”, “experimental design”, or “factorial” till 2016. The search was ended on February 6th 2017. Around 3,000 papers are found to include one or more keywords. We do not claim that the review presented here is comprehensive, because many papers are believed to use DoE without including any of the keywords used in the current search. In addition, errors might have occurred during collecting and analyzing the data from these 3,000 publications.

All the 3,000 papers are checked to decide if they actually used DoE. We found that the majority (2,019) did not implement DoE. The papers appeared in the search because one or more of the keywords are misused or are present in their literature review or references list. We found that 981 papers - which represent around 0.49% of the 200,000 total onepetro publications until 2016 - actually implemented DoE in their work. These are divided into 227, 749, and five for experimental, computer modeling, and social studies, respectively. Included in the 981 publications, there are 137 journal papers, which are divided into 42, 94, and one for experimental, modeling, and social studies, respectively. The majority of objectives fall under DoE applications of screening, metamodeling, optimization, and uncertainty analysis. **Fig. 2.2** shows the changes in the annual count and percentage of simulation and experimental DoE publications. Modeling publications started to dominate the annual count in early 2000s. The percentage was relatively stable with an average of 0.23% until 2000. Afterwards, it started to increase until 2006 when it became relatively constant again (with a slight increase over time) at an average of 0.80%. **Fig. 2.3** shows the classification of the publications among the publishers based on the

nature (modeling, experimental, and social) of the study. As expected, modeling dominates.

2.3.1 DoE in Experimental Studies

DoE has been used in petroleum engineering literature for over seven decades. The earliest reference found on SPE online library (onepetro.org) dates back to 1946, when Kemler (1946) did lab experiments to define the effects of thread tolerances, makeup, and type of lubricant on leakage resistance of threaded joints. Ten years later the next DoE experimental work was published (Krueger 1956). It used factorial design to investigate the effects of five variables on perforation performance. Vogel (1956), which is sometimes erroneously conceived as the first DoE publication in onepetro literature, extended the work of Krueger (1956) by developing a fixed-effects model to fit the experimental data. Out of a total of 227 experimental DoE publications, 33 did not mention the type of DoE they used. There are 15 repeated publications (conference papers converted to journal ones).

The total annual number of publications increased very slightly until 2010, when it started to accelerate with a total of 74 papers (68 conference and six journals) published between 2011 and 2016. However, the percentage of DoE publications does not increase with time for either conference or journal publications. More awareness about the potentials of using DoE in physical experimental is required.

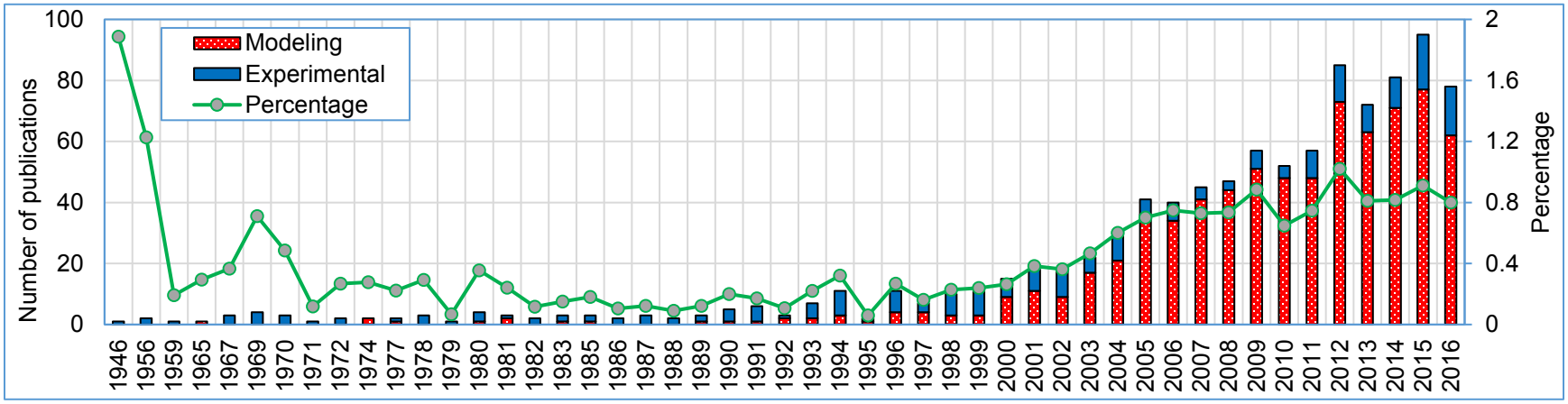
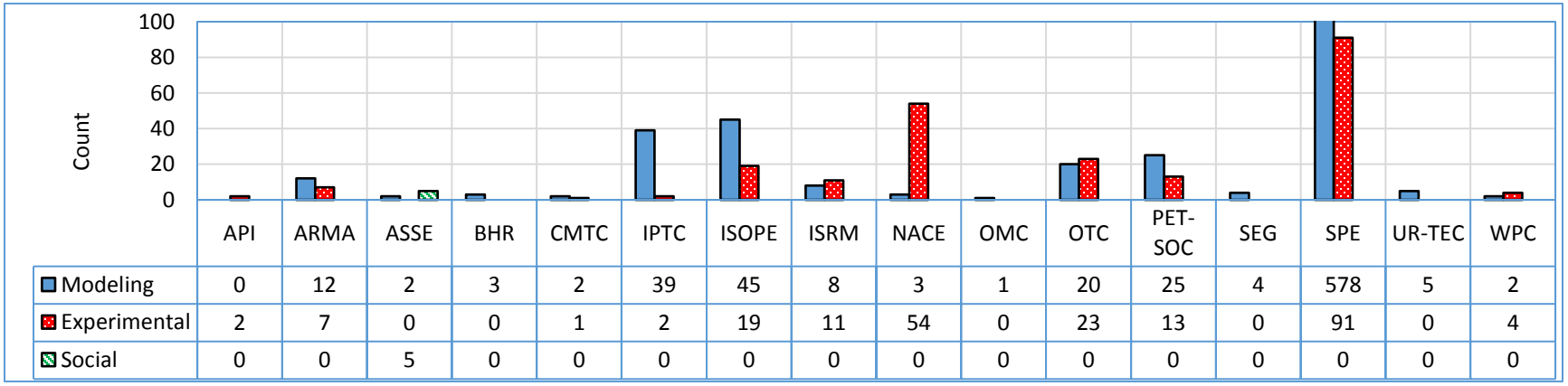


Fig. 2.2—Numbers and percentage of DoE papers.



	API	ARMA	ASSE	BHR	CMTC	IPTC	ISOPE	ISRM	NACE	OMC	OTC	PET-SOC	SEG	SPE	UR-TEC	WPC
Modeling	0	12	2	3	2	39	45	8	3	1	20	25	4	578	5	2
Experimental	2	7	0	0	1	2	19	11	54	0	23	13	0	91	0	4
Social	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 2.3—Classification of the publications among various publishers based on the nature of the study (modeling, experimental, or social). SPE modeling publications are 578. The scale ends at 100 for clarity purposes.

DoE applications in experimental studies include sensitivity and screening (64.3%), metamodeling (23.7%), optimization (11.4%), and robustness (0.6%). As expected, the major application is sensitivity. The maximum number of factors is 16 (Rosenberg and Syrett 1996) followed by 11 (Kapustka et al. 2005). The most frequent number is three. The average number of factors is almost constant with time at around four. This is expected taking into account the cost of physical experiments which limits the number of factors that can be tested in labs. Factorial designs represent 59.40% of the total DoE types reported, which is expected.

For proxy generation, multi-linear regression is the dominant technique (94.9%), which could be attributed to its easiness and flexibility. It is also appropriate for physical experiments, whose nature involves random error. Quadratic models dominate regression models at 71.2%. This is expected because, unlike linear models (23.3%), they account for nonlinearities, and they have less terms than cubic models (5.5%), which is translated into less model parameters and therefore, less experiments.

2.3.2 DoE in Modeling Studies

The first modeling DoE paper found on SPE online library (onepetro.org) is Winestock and Colpitts (1965), which used a gas-well simulator. The next one was published by Sawyer et al. (1974), which employed a fractional factorial design to study the effects of nine variables on the economics of wet gas drive and develop a simple proxy with linear and interaction terms. Like Vogel (1956), Sawyer et al. (1974) or Damsleth et al. (1992) are mistakenly considered to be the first to introduce DoE to reservoir simulation studies or to the petroleum industry in general by some onepetro publications. Until 1980, only two more modeling papers were published (Gregory 1974; Klasi 1980). The annual

count of DoE publications did not change until the 1990s when it began to slightly increase with time. From 2000 to 2016, the average annual increase was substantial with a total of 646 papers published during that period. The percentages of journal and conference publications demonstrate a similar behavior until 2009, after which they remain relatively stable at an average of 0.73% and 0.70%, respectively.

Out of a total of 749 modeling DoE publications, there are 62 repeated publications handling the same problem (conference papers converted to journal publications). In addition, 212 out of the 749 did not state which DoE technique they used. For the remaining 537 that used DoE, only 139 papers used modern DoE methods (space-filling designs) that are suitable for simulation. The remaining 398 papers only used classical DoE techniques that are conventionally used in physical experiments, and are sometimes not well-suited for computer experiments. Thus the conclusions of such publications might be questionable, if not incorrect. The space-filling designs represent 20.36% of the total DoE types reported.

DoE applications in simulation studies include sensitivity analysis and screening (40.7%), metamodeling (25.9%), uncertainty analysis (15.3%), optimization (9.7%), and history matching, a special case of optimization (8.4%). The maximum number of factors is 125 (Zheng et al. 2016). The most frequent number is five and the average is around nine (9.1), both being higher than their corresponding values in experimental studies. Contrary to experimental studies, the average number of factors in simulation studies has doubled in the last 40 years. This could be attributed to the increased computational power and speed. For proxy generation, least squares regression is the dominant technique

accounting for 71.7% of the total, with quadratic models being the most frequently used at 76.6%. Least squares regression, however, is not well-suited for computer experiments unless the regression model is checked, which is rarely the case. Other major proxy models used include kriging (15.8%), artificial neural networks (5.8%), thin-plate splines (2.2%), and radial basis functions (1.9%).

2.4 Peculiarities in Computer Experiments

We have shown that the most common way to construct a proxy using DoE in reservoir simulation is by applying classical DoE techniques (e.g., factorial and composite designs) and continuing with RSM and regression analysis to generate second-order polynomials that represent the computationally expensive reservoir simulation runs. However, the conclusions of such publications are debatable from a statistical viewpoint and the results might not be correct. The doubts arise mainly because computer experiments are deterministic in nature with no random errors. Thus statistical analyses cannot be applied appropriately (Sacks et al. 1989b; Simpson et al. 1997; Kleijnen 1998; Giunta et al. 2003). In addition, computer experiments could handle more factors, have considerably higher-order (>2) interactions, and are characterized by more complex nonlinear response-factors relationship (Sanchez 2005). Hence, this section calls attention to the common problems in applying DoE and metamodeling in computer experiments in the petroleum industry. We also provide some general recommendations regarding the appropriate use of different DoE and metamodeling methods.

2.4.1 Pitfalls in Applying DoE

Classical DoE techniques are widespread in modeling studies in petroleum engineering. They account for 79.6% of the total DoE techniques used. They were primarily designed for physical experiments, which involve random errors due to the random nature of the factors and the presence of nuisance variables. Deterministic computer experiments do not have random error. Thus applying the three basic principles of classical DoE (randomization, blocking, and replication) does not make any sense. In addition, choosing variable combinations to be on the boundaries of the variable space (which is done to strengthen the signal (effect) and comparatively minimize the effects of random errors in classical DoE) is not the best practice in computer experiments because this might produce inaccurate predictions in parts of the design-space, which are not well covered. As a result, classical designs might be inappropriate, and sometimes inefficient, for the deterministic computer experiments (Sacks et al. 1989b; Simpson et al. 2001a; Giunta et al. 2003; Santner et al. 2003; Kleijnen et al. 2005; Wang and Shan 2006; Kenett et al. 2014; SAS Institute Inc. 2015).

Additionally, relying on two-level factorial and fractional factorial designs to estimate the effects of variables might lead to misleading results because modeling response may not be monotonic in nature when one or more input variables change their levels. Moreover, since three-way and higher interactions are often neglected, fractional factorial designs might lead to the estimation of biased main and two-way interaction effects. Another important point is that classical DoEs are usually designed in such a way that minimizes the number of required experiments. This is not necessary for computer

experiments because computers can handle a considerably larger number of experiments. Computer modeling can also handle more factors (Sanchez 2005; Giunta et al. 2003; Law 2014).

Thus instead of having a small number experiments distributed around the boundaries of low-dimensional design space with replicates at the design points, the experiments should be spread out to fill the interior of the variable space (whose dimensionality can be increased by incorporating more factors than classical DoEs can handle), as well as its edges with no replicates. The design points can be either uniformly distributed throughout the design space or are spread out as far from each other as possible. This ensures providing information about all parts in the design space. In addition, this reduces the bias (difference between fitted proxy and modeling output) in response prediction, which is often based on interpolation. This is because the error of prediction depends on the distance between the point of interest and the closest design points. Such "space-filling" designs constitute modern DoE. Modern DoE is sometimes called DACE, or design and analysis of computer experiments. However, DACE is also often used to mean kriging (Sacks et al. 1989b; Simpson et al. 2001a, 2001c; Giunta et al. 2003; Santner et al. 2003; Kleijnen et al. 2005; Chen et al. 2006; Wang and Shan 2006; Kenett et al. 2014; SAS Institute Inc. 2015).

Space-filling designs are more flexible than classical designs, as they do not require making prior assumptions regarding the nature of the relationship between the response and input variables. Therefore, they help avoid the bias associated with such assumptions and allow fitting a variety of models (Sacks et al. 1989b; Simpson et al. 2001a; Giunta et

al. 2003; Santner et al. 2003; Kleijnen et al. 2005; Kenett et al. 2014; SAS Institute Inc. 2015). Another feature in which modern DoE is superior to classical DoE is that it can handle probability distribution of the factors. Classical designs assume uniform distribution. Although this might be a good assumption in physical experiments, it is not generally the case in computer modeling (Giunta et al. 2003).

There are various sampling strategies involved in space-filling designs. Latin hypercube design (LHD) is one of the most common space-filling designs (Kenett et al. 2014), accounting for 72.2% of the total space-filling designs used in SPE publications. It was first introduced by McKay et al. (1979) and has recently been extensively used in reservoir simulation studies that apply DoE. Other designs that can be used in computer experiments include sphere packing, uniform, minimum potential, maximum entropy, Gaussian process IMSE optimal, pseudo-Monte Carlo sampling, quasi-Monte Carlo sampling, Hammersley sequence sampling, nearly orthogonal designs, orthogonal arrays, and fast flexible filling (Simpson et al. 2001a; Santner et al. 2003; Chen et al. 2006; Bingham et al. 2009; SAS Institute Inc. 2015). The average percentage of modern DoE in the last six years is 31.1%, which is promising considering its continuously increasing implementation with time.

2.4.2 Pitfalls in Applying Metamodeling

Least-squares regression technique is most common in modeling studies since it is well-established and simple. Using least squares to fit a mathematical equation to the results of physical experiments makes sense because the random results affected by noise are averaged. In deterministic reservoir simulation experiments however, there is no

random error. Thus there is no point in averaging the response at the design points (Simpson et al. 1997). In addition, the error of approximation could not be used to account for random errors. This means that the statistical analysis and ANOVA done on least squares regression in modeling are not valid.

To use regression proxies, at least verification and validation (V&V) should be done to ensure its accuracy. Validation of a proxy is a measure of its accurate representation of the real simulation model results (Carson 2002). It is can be done using blind random test data at non-design points (Simpson et al. 1997, 2001c). If such data is not available, cross-validation might be a good option (Schneider and Moore 2000). If cross-validation is not enough to assess model accuracy (Lin 2004), other measures of fit like determination coefficient (R^2), maximum absolute error, root mean square error (RMSE), and predicted residual error sum of squares (PRESS) can be used to indicate the quality of fit because they are less sensitive to the absence of randomness (Simpson et al. 1997, 2001c; Wang and Shan 2006; Myers et al. 2009). However, such measures are usually not enough (a high R^2 does not necessarily indicate a good fit) and verification should be done. Verification of a proxy is to check whether it conforms to specifications and assumptions (Carson 2002). Verification of regression proxy adequacy can be applied using regression diagnostics (Sheather 2009) to check assumptions like constant variance and normality of residuals (Simpson et al. 1997, 2001c). This is very important if the model is used mainly for inference (as opposed to prediction).

Unfortunately, proxy verification and validation are not widely applied. For example, a very few DoE publications were found to deal with residual diagnostics

(Cobianco et al. 1999; Castillo et al. 2010; Al-Mudhafar and Rao 2015). A quick search for the keyword “regression” on onepetro reveals that it appeared 11,767 times till 2016. On the other hand, the keywords “regression diagnostics”, “regression diagnostic”, “residual diagnostics”, “residual diagnostic”, “residual analysis”, and “residuals analysis” appeared only 47 times combined. This implies that regression diagnostics are rarely applied. Therefore, the predictions made by the least-squares equations may be questionable, if not incorrect. This brings into question the reliability of the results of analyses based on regression models. However, it is important to note that there is no reason to reject a least squares RS when it provides a good match between predicted and actual responses (Simpson et al. 1997; Law 2014). This is especially true if the model is used mainly for prediction (as opposed to inference).

Another concern with least squares regression is that its models may yield poor predictive accuracy in stiff non-linear problems or when the investigated variables’ space is very large (Simpson et al. 2001c; Li and Friedmann 2005a, 2005b; Kleijnen 2010; Osterloh et al. 2013; Law 2014). Computer experiments are known to have more complex nonlinear response-factors relationship (Sanchez 2005). Higher-order polynomials can improve the proxy’s predictive accuracy. However, they need lots of design points to estimate a larger number of regression coefficients and they might be unstable. Other metamodeling techniques could solve or minimize the problems associated with nonlinearity and the absence of random errors.

Artificial neural networks (ANN) can produce accurate proxies despite requiring intensive training and large data sets, which is computationally costly in terms of modeling

time and effort. However, like least squares regression, the accuracy of ANN is reduced because it is not data exact as it smooths out the predictions. Despite being appropriate for physical experiments, this averaging might not be acceptable for computer experiments where random error is not existent. Interpolation techniques (e.g. kriging (KG) and thin-plate splines (TPS) interpolation) can take care of this problem because they exactly honor the design points (Simpson et al. 2001c; Santner et al. 2003; Li and Friedmann 2005a, 2005b; Zubarev 2009; Osterloh et al. 2013). Moreover, they could enhance the computational efficiency. KG is promising, as it has accurate predictability (Sacks et al. 1989b). In addition, KG (or nonparametric Gaussian process) outperforms TPS interpolation for higher nonlinearity despite being more complex and computationally costly (Zubarev 2009). Zubarev (2009) recommended TPS interpolation as the most reasonable solution in spite of being more prone to error for small data sets. Nevertheless, both interpolation methods tend to smooth out the non-linearities and are less efficient in case of unevenly distributed nonlinearities (Li and Friedmann 2005a, 2005b). Least squares regression is increasingly replaced by other techniques, especially kriging whose average percentage in the last five years is 20.9%.

2.5 Criticisms against DoE and Metamodeling

This section provides an overview of the criticisms against DoE and metamodeling and some recommendations regarding these criticisms.

2.5.1 DoE

One of the disadvantages of DoE found in the literature is that most DoE do not honor the probability distribution of the factors. Li et al. (2011) compared the probabilistic

performance of the proxies of classical (D-optimal, exhaustive sampling, and folded Plackett–Burman) and space-filling designs with that of probabilistic collocation method (PCM), which can handle the probability density functions of the factors. They concluded that PCM is more efficient and accurate than DoE. However, space-filling designs can account for the probability-density function of the input variable (Giunta et al. 2003), which – if considered – might lead to different conclusions. For classical designs - which assume uniform distribution - different weights could be assigned to the different design points according to their joint probability while generating the proxy. Some more work needs to be done in these two respects. According to Wolff (2010), however, the distributions of uncertain variables may have small influence on forecasts as compared to the larger influence of their ranges. This assumption needs to be supported, especially when using classical DoEs.

Another disadvantage is associated with high dimensionality (say 100) in computer modeling. Breaking down the complex problem into many smaller problems, with smaller number of variables, is one way to solve this issue. This technique was used in other disciplines (more details in Simpson et al. 2001c). For screening purposes, classical screening designs (like fractional factorial and Plackett–Burman designs) would require too much time because at least (k + 1) scenarios should be run. Special screening designs like group screening, iterated fractional factorial designs, Morris's designs, sequential bifurcation, supersaturated designs, and Trocine screening can handle such situations (Bettonvil and Kleijnen 1997; Lewis and Dean 2001; Trocine and Malone 2001; Morris 2006; Wan et al. 2006; Kleijnen 2008; Kleijnen 2010). However, to the best of our

knowledge, such methods have not been applied in petroleum engineering. This could be because only three publications are found to deal with more than 50 factors; Agrawal et al. (2015), Schulze-Riegert et al. (2016) and Zheng et al. (2016) used 70, 116, and 125 factors, respectively. The first two applied screening using LHD and OFAT, respectively, and the latter used all the factors to generate a proxy using ANN.

One more problem that might be combined with the high dimensionality issue is the fact that DoE assumes that the factors are independent. This is not always a problem, however. For example, Egeland et al. (1992) has shown that the dependence of variables has a little effect on Monte Carlo simulation. This is not generally the case though. If dependence is detected, its effect should be investigated. One way to address this problem is to use principal component analysis, or PCA (Wang and White 2002). PCA can reduce the number of variables depending on their correlations with each other. In their study, Wang and White (2002) reduced the number of factors from 36 geologic factors to six independent groups of factors, which is manageable by DoE and satisfies the assumption of independence.

Two other major concerns in DoE are adding factors that were initially overlooked in the middle of the study and dealing with factors with different number of levels. Optimal designs can take care of these two problems (White and Royer 2003). Although optimal designs could be appropriate to apply in physical experiments, care should be exercised when they are applied to deterministic computer experiments because of the lack of random error.

OFAT might be preferred over DoE when the experiments are expensive. However, sequential DoE is another alternative. This technique has been studied extensively in classical DoE. However, a lot of work needs to be done to evaluate its potentials in computer modeling in various disciplines (Lin et al. 2004; Chen et al. 2006). Petroleum engineering is not an exception.

2.5.2 Metamodeling

The major concern about applying the traditional RSM is the absence of random errors in deterministic computer experiments. However, if randomness could be introduced into the model (in one or more factors), inferential statistical analysis would make sense. If not feasible, then the model should at least be verified using regression diagnostics (particularly if the model is used for inference), validated at untested points, and/or cross-validated. Inferential statistical analysis should be used with caution, and test statistics (F, t, and others) should be considered as descriptive rather than inferential statistics (Yamada 2008).

One of the disadvantages of using proxies is that they might not consider the important physical effects and could be inefficient in locating the optima, especially when nonlinearity is severe and unevenly distributed (Ludvigsen and Le 2015). To mitigate such problems and improve the accuracy, Li and Friedmann (2005a) proposed applying interpolation techniques to partitioned parts of the variable space. However, this could be computationally intensive. Thus Li and Friedmann (2005b) introduced a novel metamodeling technique (TPS interpolation based on amplitude factor analysis) that can handle nonlinearity without partitioning the variable space. This enhances the

computational efficiency. They claim that it is more efficient and accurate than TPS interpolation and regression.

According to Lawal (2009), metamodeling has many shortcomings. One of the major concerns is that the input distributions are not chosen carefully. For example, engineers commonly use P10 and P90 for factors to generate a proxy, which is then used to estimate the whole range of uncertainty, including P0 and P100. However, this is not a problem of the proxy itself. Rather, it is related to how the proxy is generated. Another drawback is that proxies do not preserve the physical relationship between various responses such that the proxies are mathematically consistent. However, this is not a drawback because no study was found in the literature to violate such consistency. One more issue is that proxies are built using modeling results, not actual reservoir performance, which could make them inaccurate. In spite of being true, the major part of this problem is related to the accuracy of the computer model itself. It is not just a problem of the proxy. It is easier to control the proxy and ensure its accuracy than it is for simulation. The non-uniqueness of the simulation model that satisfies the history matching is a problem of the simulation, not the proxy. Proxies can be viewed as lower-risk approximations of higher-risk approximations. So if the model is not well-tuned, the proxy will simply follow the "garbage in, garbage out" principle. On the other hand, if the simulation model well describes the physical process, a validated and verified proxy would give good approximations of reality. Another problem that Lawal (2009) pointed to is that the inclusion of controllable design variables (e.g., fracture half-length) in a proxy with uncontrollable intrinsic reservoir properties (e.g., permeability) might mask the effects of

the latter. Close attention should be paid to this point. It might be of interest though, to see which of the controllable or uncontrollable variables have the dominant effect. This could have very important implications in decision-making process. For example, if the uncontrollable variables are overwhelming, the priority should be given to searching for sweet reservoir spots rather than optimizing the design variables.

The author also disagrees with many of the other "shortcomings" provided in Lawal (2009). First, derivatives and integrals of proxies with respect to time are invalid unless the proxy includes time as a variable. For instance, proxies for total oil production and oil flow rate give their respective responses at a certain point of time. Thus taking the time derivative of the first to calculate oil flow rate or integrating the second to calculate the total oil production is not justifiable. Another concern about the shortcomings that Lawal (2009) mentioned is that he used the proxy for extrapolating the performance outside the investigated range. For example, a proxy for recovery factor (RF) as a function of permeability (in addition to some other variables) was used to predict RF in case of zero permeability rock to get a non-zero RF. However, proxies cannot be used outside the investigated range of the factors. Thus this point is invalid as well.

Box (1979) said, "All models are wrong but some are useful." No model can capture the exact physical nature of relationship between the variables and responses. Zubarev (2009) recommended understanding the proxy limitations and performing proper quality assurance to quantify the errors before taking decisions. In spite of their imperfection, generating proxies is still viewed as an efficient and valid tool for the reasons discussed previously in DoE Applications (section 2.2.1).

2.5.3 Choice of Sample Size

For physical experiments, where regression analysis is a legitimate metamodeling technique, the recommended number of required experiments per regression coefficient in the regression model ranges between 5 and 20 (Schmidt 1971; Green 1991; Harrell 2001) with 20 being preferred (Green 1991). So for a full quadratic polynomial with two factors (six regression coefficients), a good number of experiments to start with could be 30 – 120 experiments.

A common issue associated with applying DoE in computer modeling is that the number of runs is arbitrarily chosen. This is true in the vast majority (more than 95%) of the petroleum engineering studies. Thus some general guidelines are required in this respect. Despite being important, not much could be found in the literature regarding the choice of sample size in deterministic computer modeling (Loeppky et al. 2009). This is especially true for large-scale problems where the sample size increases exponentially with the number of variables, a problem known as the "curse of dimensionality" (Wang and Shan 2006).

Although the true relationship between the response and the variables is usually unknown, the nature of the model could sometimes be assumed. In doing so, one can use some criteria for choosing a design. This could help facilitate selecting the sample size. For example, Osterloh et al. (2013) concluded that to get a good predictive accuracy of an LHD-based metamodel, the sample size should be 1 to 1.5 times the number of coefficients in a second order polynomial ($k = (n+1)(n+2)/2$; n is the number of variables). Kaufman et al. (1996) found that samples of size 1.5k for lower dimensional problems (5 – 10

variables) and 4.5k for higher dimensional problems (20 – 30 variables) are required to get good accuracy of second-order polynomial models.

Chapman et al. (1994) and Jones et al. (1998) suggested using a sample size 10 times the dimension of the variable space (10d). This rule of thumb however might not be adequate as dimensionality increases because it leads to a sparse distribution of design points. Loeppky et al. (2009) studied this informal rule and concluded that 10d is adequate for an initial experiment when d is five at most. The rule also works with d up to 20 or more if the response is sensitive to a few of the investigated factors. In general, they recommended checking the proxy and adding more design points if it lacks accuracy. An example of the inadequacy of the 10d rule was given in Chen et al. (2010). They used computer simulation to model bistable laser diodes with two factors (d = 2). They found that the kriging proxy was not quite accurate over the entire space and, therefore, more than 10d (20) design points are required.

Focusing on space-filling designs, Simpson et al. (2001a) found that larger sizes generally improve the accuracy. They also found that larger sizes do not affect the predictive accuracy of the metamodels (except for MARS) when non-linearity is not severe. On the other hand, Liu (2005) found out that, excluding the extremely nonuniform space-filling designs, the sample size is mostly more influential on the predictive accuracy of the metamodel than the type of design used. The distribution of the design points is less important as long as the best sample size is unknown (Wang and Shan 2006). This agrees with Johnson et al. (2011) who concluded that using an adequate sample size of space-filling design is the best way to enhance the predictive accuracy of Gaussian-process

models regardless of the type of space-filling design used. The "adequate" size depends on the complexity of the relationship between the response and its factors and, therefore, is difficult to know a priori and will remain a mystery (Wang and Shan 2006; Johnson et al. 2011). That is why sequential and adaptive sampling is becoming more popular (Wang and Shan 2006).

Sequential sampling was generally found to be more efficient than one-stage sampling and is less computationally demanding (Jin et al. 2002). In addition, it can help determine an interesting region in the design space where sampling can further focus on to improve the metamodel prediction accuracy. However, they also concluded that sequential sampling with or without adaptation is not guaranteed to improve the accuracy of a metamodel over one-stage sampling. Lin (2004) developed sequential exploratory experimental design (SEED) using D-optimal and maximum entropy sampling to increase metamodel accuracy. Sample points are generated using the analysis of information from data/validation points and metamodels. Sasena et al. (2002) used a global constrained nonlinear optimization algorithm (superEGO) to create adaptive DoEs whose sample points are selected based on information acquired from previous samples.

Jin et al. (2001) recommended further investigations in adaptive sampling, where the variables are sampled based on their contribution to response. Adaptive response surface method (ARSM) was developed by Wang et al. (2001) based on central composite design (CCD). For each step, ARSM uses a critical value to reject regions of the design space whose response surpasses such a limit. This strategy gets the design space closer to the optimum region. This technique, however, requires a sample size that increases

exponentially with the number of variables. In addition, the points of previous steps cannot be inherited to the new sample. Thus Wang (2003) replaced CCD with an inheritable Latin Hypercube design (LHD) to avoid these two limitations. It is claimed that the new technique is better suited for global optimization of computationally intensive problems.

Still however, more research is required to address large-scale problems and new sampling and metamodeling techniques are required (Wang and Shan 2006).

2.6 Comparison between Metamodeling and DoE Techniques

Many authors compared different proxy and DoE techniques. Appendix A summarizes some of the work done in petroleum engineering literature. There is a general agreement in petroleum engineering literature that space-filling designs are better than classical designs. This agrees with the findings of Simpson et al. (2001c) and Giunta et al. (2003). In addition, Johnson et al. (2011) found that no one space-filling design has an advantage over another regarding prediction accuracy using Gaussian process models. For metamodeling, however, there is no consensus regarding which technique is superior to the others, which agrees with Simpson et al. (2001c) and Wang and Shan (2006). Chen et al. (2006) provided detailed comparison tables for DoE and metamodeling techniques.

Chen et al. (2006) recommended developing DoEs and metamodeling techniques that are better suited for different applications. General guidelines and recommendations about which DoE and metamodeling techniques (especially for computer modeling) are not as well-developed in petroleum engineering literature as in other disciplines. Petroleum engineers and scientists can use those guidelines developed in other disciplines (Appendix A) until more extensive studies are done (Simpson et al. 1997; Giunta and Watson 1998;

Simpson 1998; Varadarajan, et al. 2000; Jin et al. 2001; Simpson et al. 2001a, 2001b, 2001c; Giunta et al. 2003; Chen et al. 2006; Wang and Shan 2006; Johnson et al. 2011). However, before doing this, other challenges need to be addressed first.

At the end of his paper, Vogel (1956) wrote "It is hoped this paper will encourage others to investigate the many possible uses of these techniques in the fields of reservoir engineering and research." Now, 62 years after Vogel's publication, it is hard to say that his hope has become realistic. Engineers and scientists fear statistics and therefore, they resist change (Tanco et al. 2010). They rarely use statistics (Box 2001). In fact, some of them think that OFAT represents a "sound" engineering principle or is related to the scientific method (Montgomery 2012). Thus it is hard to realize that there is a problem in the first place. Petroleum engineers and scientists are no exception, and OFAT and best-guess approaches are ubiquitous in the petroleum industry. Part of the problem could be the lack of DoE awareness, which could be attributed to not focusing on DoE and statistical analysis in general in petroleum engineering programs. We need to focus on statistics as an indispensable way to solve engineering problems. Leveraging the enormous potentials of statistics and DoE would help us improve our approaches and increase the efficiency of various operations in the petroleum industry.

2.7 Conclusions

Design of Experiments (DoE) is the gold standard systematic research methodology that can maximize analytical research efficiency by efficient planning and execution of experiments and unbiased analysis of the results to establish cause-effect relationships. We recommend adopting DoE as a standard research methodology in the

petroleum industry. This could be started by incorporating it as one of the crucial courses in petroleum engineering programs. On the long run, we - petroleum engineers - can draw our own guidelines for the best choice of sample size and the best use of DoE and metamodeling in various situations, which could enrich the industry and maximize the efficiency of its various processes. We hope that this work could motivate petroleum researchers and engineers to learn and apply statistics and DoE, regardless of the type of experiments that they conduct; physical experiments, modeling runs, or field trials.

3. OPTIMIZATION OF PRIMARY AND ENHANCED OIL RECOVERY IN EAGLE FORD SHALE USING STATISTICAL DESIGN OF EXPERIMENTS

3.1 Introduction

In this chapter a DoE-based workflow is developed and applied with one major goal: to determine the potential of optimizing four recovery schemes to maximize the low RF in Eagle Ford shale. The four schemes are primary production (PP), waterflooding (WF), continuous miscible gas flooding (CON), and huff and puff miscible gas injection (HNP). The focus in this chapter is on controllable design variables (e.g., fracture spacing). Uncontrollable reservoir state variables will be considered in Chapter 4. That major goal is done by achieving the following objectives:

1. Estimate the effects of design variables on RF
2. Define the most important variables that affect the RF of PP, WF, CON, and HNP
3. Specify if PP optimization is enough to improve the low primary RF
4. Compare the four recovery schemes to identify the most prospective one
5. Determine which miscible injection scheme (CON vs. HNP) is better
6. Specify the injection gas that yields the best economic performance
7. Determine the optimum combination of important design variables that maximizes the RF for selected recovery schemes
8. Evaluate the potential of an innovative injection scheme that uses alternating injecting and producing fractures along the same lateral

3.2 Methods

3.2.1 DoE

JMP and R statistical software are used here for statistical design and analysis. The DoE workflow starts with selecting RF as a response that needs to be maximized by optimizing the effects of controllable design variables. The next main steps (**Fig. 3.1**) represent three of the major applications of DoE (Eriksson 2008; NIST/SEMATECH 2016). These steps are as follows:

1. Screening: The uncertain design variables are varied based on a screening DoE and simulation is run to specify the most important variables that affect RF. The changes in the rank of importance of the important variables is also tracked through a production period of 15 years.
2. Optimizing: The optimum settings of the most important variables that maximize the RF are determined.
3. Comparison: RFs of all recovery schemes are statistically compared together to determine which one is more prospective.

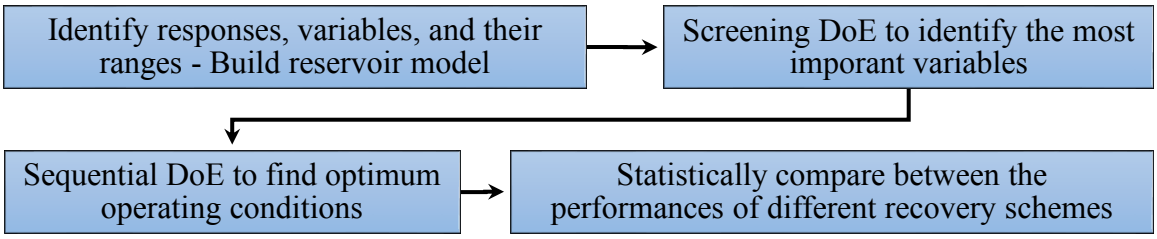


Fig. 3.1—Flowchart of the DoE-based optimization process.

More details on how these steps are applied are given in section 3.3.

3.2.2 Simulation Model Description

Simulation is one of the most widely accepted approaches used to evaluate reservoir performance and forecast production. It is a strong managerial tool to optimize oil recovery before lab testing or field application. **Table 3.1** summarizes the properties of reservoir rock and fluids (Chaudhary 2011; Morsy et al. 2013; Wan et al. 2013b; Gamadi et al. 2014; Wan et al. 2014; Fragoso et al. 2015; Wan et al. 2015) of the single-porosity model used in this study. CMG is used for modeling.

Property	Unit	Value
Depth to top of the reservoir	ft	10,500
Pressure	psi	7,350
Temperature	°F	225
Thickness	ft	200
Porosity	%	7
Permeability	nd	500
Vertical-to-horizontal permeability ratio		0.1 (1)†
Rock compressibility	psi^{-1}	5.00E-06
Irreducible water saturation (S_{wi})		0.3 (0.01)†
Residual oil saturation (S_{or})		0.3 (0.01)†
Critical gas saturation		0.05 (0.01)†
Endpoint relative permeability to oil at S_{wi}		1 (1)†
Endpoint relative permeability to water at S_{or}		1 (1)†
Endpoint relative permeability to gas at immovable liquid saturation		1 (1)†
Oil exponent		3 (1)†
Water exponent		3 (1)†
Gas exponent		3 (1)†
Liquid exponent		3 (1)†
Gas gravity		0.8
Oil API gravity	°API	42
Bubble point pressure	psi	2,375
Gas-oil ratio	scf/STB	653
† Values between parentheses are for fracture blocks.		

Table 3.1—Properties of reservoir rock and fluids for the present simulation model.

Horizontal wells are simulated in the middle of the thickness of the formation. Transverse fractures (TFs) are assumed to be perpendicular to the direction of the horizontal well and equally spaced with heights equal to the formation thickness. **Fig. 3.2** illustrates a part of two wells with multistage hydraulic fractures. Because of symmetry, only a small part representing fracture wing is simulated (Fig. 3.2b). Production is simulated for 15 years.

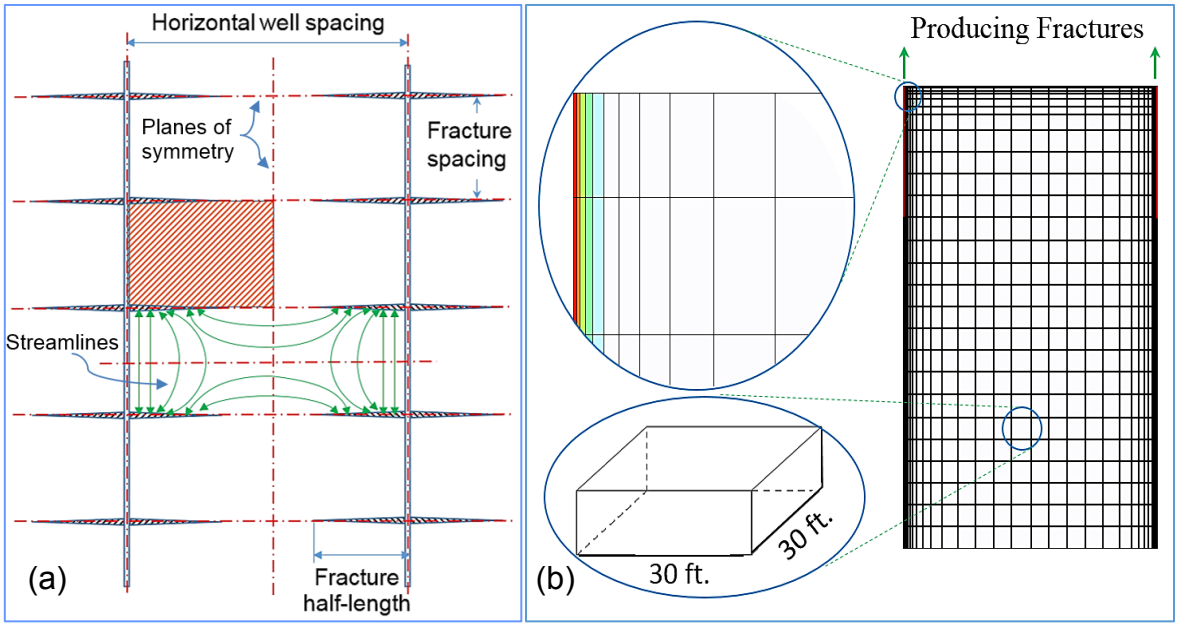


Fig. 3.2—(a) Schematic illustration of a part of two wells with multistage hydraulic fractures; (b) Enlarged simulated region showing LGR, permeability gradient away from hydraulic fracture, and non-refined block dimensions.

The relative permeability curves of the fracture and the matrix are generated using modified Brooks-Corey functions. The reservoir is assumed to be homogeneous with no natural fractures. The actual width (w_f) and porosity (φ_f) of the fractures are assumed to be 0.0025 ft (0.03 in) and 0.4 (porosity of orthorhombic proppant packing), respectively. The

simulated effective fracture width, w_{eff}, is determined to be 0.1 ft because it is the minimum value found to be numerically stable. For the same reason, local grid refinement (LGR) is used around TFs and the producing perforation in the lateral and vertical directions (Fig. 3.2b). LGR also helps obtain a better representation of the rapid pressure and saturation changes around the fractures. A non-refined grid block has dimensions of 30 ft x 30 ft in the lateral directions. Fracture blocks have an effective permeability, k_{eff}, calculated using the following equation:

$$k_{eff} = \frac{k_f \times w_f}{w_{eff}} \qquad \text{(Eq. 3.1)}$$

where k_f is the actual fracture permeability.

To obtain more accurate results, the effective porosity (φ_{eff}) of fracture blocks is modified using the volumetric average of matrix porosity (φ_{mat} = 7% as shown in Table 3.1) and fracture porosity (φ_f) using the following equation:

$$\varphi_{eff} = \varphi_{mat} + \varphi_f \times \frac{w_f}{w_{eff}} = 0.07 + 0.4 \times \frac{0.0025}{0.1} = 0.07 + 0.01 = 0.08 \qquad \text{(Eq. 3.2)}$$

To avoid large transmissibility contrasts and to get a well-conditioned model that better represents the secondary fractures and microfractures induced by hydraulic fracturing treatment in the vicinity of the main fractures, permeability is reduced logarithmically in the LGR zone away from the main fractures, from k_{eff} to matrix permeability (Fig. 3.2). The width of that zone of secondary fractures increases with fracture conductivity. For example, its width is 3.57 ft if conductivity is 1,000 md.ft and 1.03 ft if conductivity is 1.0 md.ft. Such widths depend on the LGR system chosen.

Vertical to horizontal permeability ratio is assumed equal to unity in all blocks with non-matrix permeability.

A literature review was conducted to identify design variables and their ranges for the four recovery schemes (**Table 3.2**) in Eagle Ford shale (Shoaib and Hoffman 2009; Chaudhary 2011; Dong and Hoffman 2013; Morsy et al. 2013; Wan et al. 2013a, 2013b; Xu and Hoffman 2013; Eshkalak et al. 2014; Gamadi et al. 2014; Han and Gu 2014; Liu et al. 2014; Sheng and Chen 2014; Wan et al. 2014; Fragoso et al. 2015; Kalra and Wu 2015; Pu and Li 2015; Rivera et al. 2015; Wan et al. 2015; Yu et al. 2014; Zhu et al. 2017). The number of factors are five, nine, 11, and 11 for PP, WF, CON, and HNP, respectively.

Case	Variable	Unit	Levels	
			Low	High
All	Frac spacing (Frac_spcg)	ft	210	1,050
	Half well spacing (Half_W_spcg)	ft	540	1,320
	Producing BHP (Prod_BHP)	psi	2,400	3,600
PP and HNP	Frac half length (Frac_half_L)	ft	120	540
	Log of frac conductivity (log(Frac_cond))		0 (1 md.ft)	3 (1,000 md.ft)
WF and CON	Production frac half length (Prod_frac_half_L)	ft	120	540
	Injection frac half length (Inj_frac_half_L)	ft	120	540
	Log of prod. frac conductivity (log(Prod_frac_cond))		0 (1 md.ft)	3 (1,000 md.ft)
	Log of inj. frac conductivity (log(Inj_frac_cond))		0 (1 md.ft)	3 (1,000 md.ft)
WF, CON, and HNP	Injection bottom-hole pressure (Inj_BHP)	psi	7,500	11,000
	Prod. time after which injection starts (Inj_time)	years	1	9
CON and HNP	Degree of miscibility (w)		0.2	0.8
	Minimum miscibility pressure (MMP)	psi	2,500 (CO_2 or Sep. gas)	5,000 (Lean gas)
HNP	Injection time (Inj_duration)	days	30	150
	Production time (Prod_duration)	days	30	270

Table 3.2—Controllable design variables and their ranges.

Factor ranges are chosen carefully to avoid dominance of numerical error on the response and the risk of overlooking a meaningful factor. Due to its wide range, fracture conductivity is expressed logarithmically. Producing bottom-hole pressure is always kept above the bubble point pressure of 2,375 psi to avoid the evolution of gas, which would adversely affect oil production (and substantially increase the simulation time). Minimum miscibility pressure (MMP) has a range that could cover the two extremes: rich separator gas or CO_2, which requires a low MMP (2,500 psi), and lean gas, which needs higher MMP (5,000 psi) to attain miscibility. Inj_time is the period of primary oil production after which injection starts.

Continuous injection into the ultra-low permeability shale does not seem to be practical. For the injection pressure wave to propagate, either the matrix permeability must be increased or the distance between the injecting and producing wells must be decreased. The former option is impossible. In addition, the second option necessitates drilling more wells, which might not be cost-effective. To solve such a dilemma, the model is developed such that when injection starts, every other fracture along the same lateral is converted from production to injection for CON and WF. This means that the same well is used for both production and injection for continuous flooding schemes (WF and CON) as well as cyclic injection (HNP). This reduces the distance required for pressure propagation between injectors and producers and does not require drilling as many laterals. This is not currently feasible. However, it is believed that it would be better than using separate wells for injection and production in these ultra-low permeability formations. Although the author authentically developed this technique, it was found out that this injection technique

was first suggested by Dombrowski et al. (2015). Zhu et al. (2017) applied CON using this technique and found that it gave good results in terms of RF improvement. In spite of the general preference of HNP over CON in the literature, the author believes that CON will outperform HNP with the help of this innovative completion technique.

3.3 Results

3.3.1 Screening (Sensitivity)

The screening study is concerned with identifying the most important variables affecting the response (15-year RF) so that the next phase of optimization would be more focused on those critical variables. The main effects and two-way interactions are the major focus here. Therefore, only the high and low levels of the variables are of interest (Table 3.2). Full factorial design is used for PP due to the small number of its factors. Fractional factorial is used for the other three cases to reduce the number of required runs. This is summarized in **Table 3.3**. It should be recognized that the specific results presented in this dissertation are valid only for the input data shown in Table 3.1 and Table 3.2. The conclusions should be valid for other sets of input that are not much different than the data used in this work. The main purpose of this research is to illustrate the use of Design of Experiments in the oil and gas industry.

Fig. 3.3 gives a box-plot comparing the 15-year RFs of the four recovery schemes. It is obvious that CON has the highest potential for RF improvement. This agrees with the previous expectations that CON could outperform HNP, as well as other schemes. Because of that (and also because HNP takes substantially more simulation time that the other recovery schemes), HNP is not considered in the optimization stage.

Input		Approach					
		Screening			Optimization		
Scheme	# factors	DoE	# runs	# important factors	DoE	# factors	# runs
PP	5	Full factorial	32	5	I-optimal	5	100 & 25
WF	9	Fractional factorial	128 (V resolution)	8	Full factorial	5	125
CON	11	Fractional factorial	128 (V resolution)	8	Full factorial	4	144
HNP	11	Fractional factorial	128 (IV resolution)	8	-	-	-

Table 3.3—Work summary. [-] means negative effect.

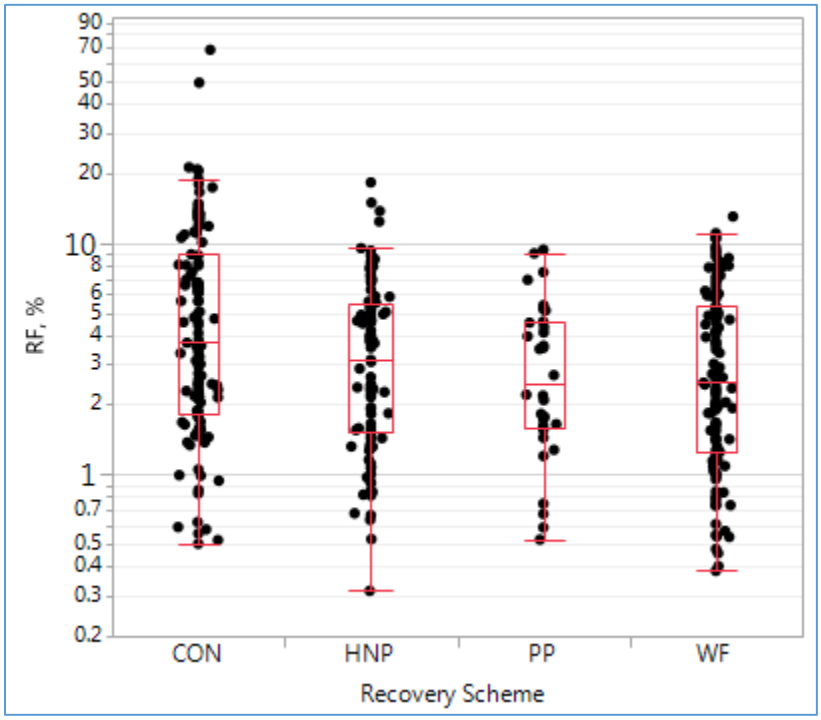


Fig. 3.3—Box-plots showing RF distribution of the four recovery schemes based on screening runs.

Pareto plot (**Fig. 3.4**) for PP shows the most important variables that affect the 15-year RF value based on a significance level, α, of 0.05[1]. All the variables of PP are important and will be carried through to the optimization stage. PP is the only one whose Pareto plot is given because the other three recovery schemes have 30+ significant variables and interactions, which would render their plots too busy and unclear. The four most important variables for all schemes are provided in **Table 3.4**.

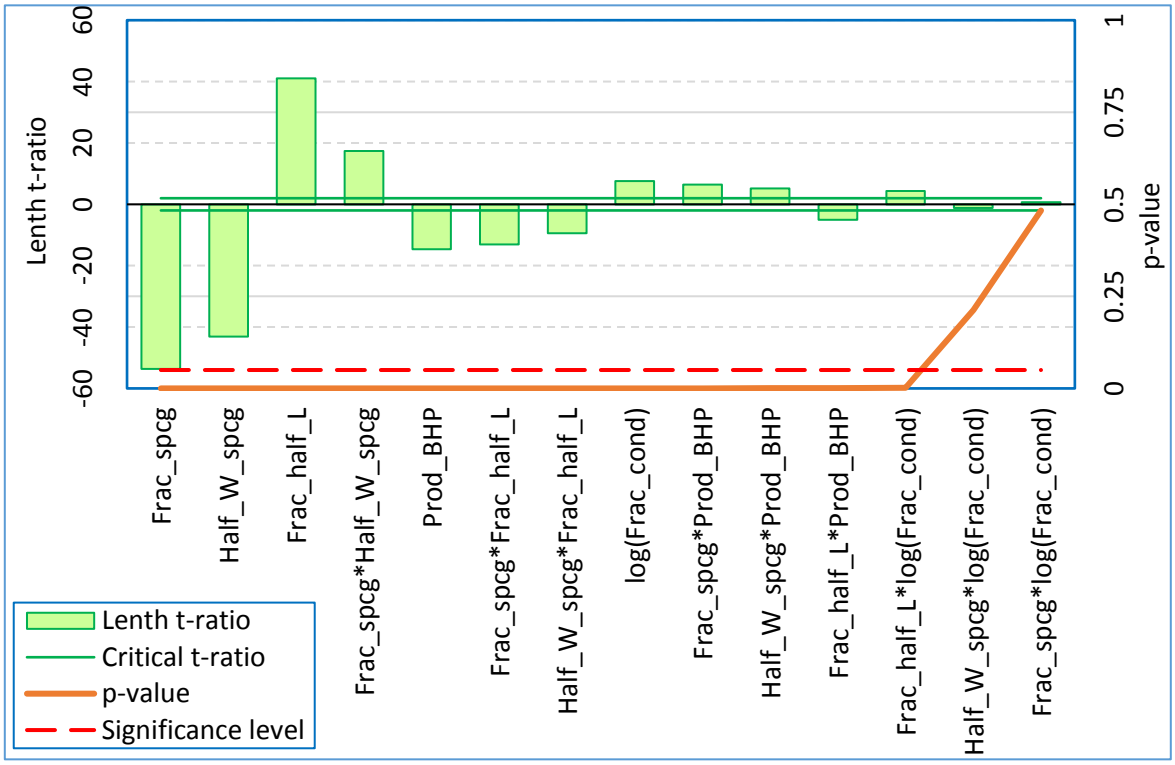


Fig. 3.4—Pareto plot for PP.

[1] "Significance" level here is only used as a value to help determine the most influential variables and is considered as a descriptive rather than an inferential statistic because of the deterministic nature of simulation.

Scheme	Most important factors
PP	Fracture spacing [-], well spacing [-], fracture half-length, fracture spacing-well spacing interaction, and producing BHP [-].
WF	Fracture spacing [-],well-spacing [-], producing fracture half-length, fracture spacing-well spacing interaction, fracture spacing-fracture half-length interaction [-], and injecting fracture half-length
CON	Fracture spacing [-], well spacing [-], producing fracture half-length, fracture spacing-well spacing interaction, fracture spacing-producing fracture half-length interaction [-], and injecting fracture half-length.
HNP	Fracture spacing [-], well spacing [-], fracture half-length, and producing BHP [-].

Table 3.4—Most important variables for four recovery schemes.

Fig. 3.5 tracks how the ranks of the most important variables change with time for PP based on their p-value. It shows that fracture spacing, well spacing, and fracture half-length dominate the influence on RF. From practical importance perspective, this is also clear in **Fig. 3.6**. This means that fracture spacing should be the variable of primary interest in the optimization stage. This behavior is also observed in all three injection schemes.

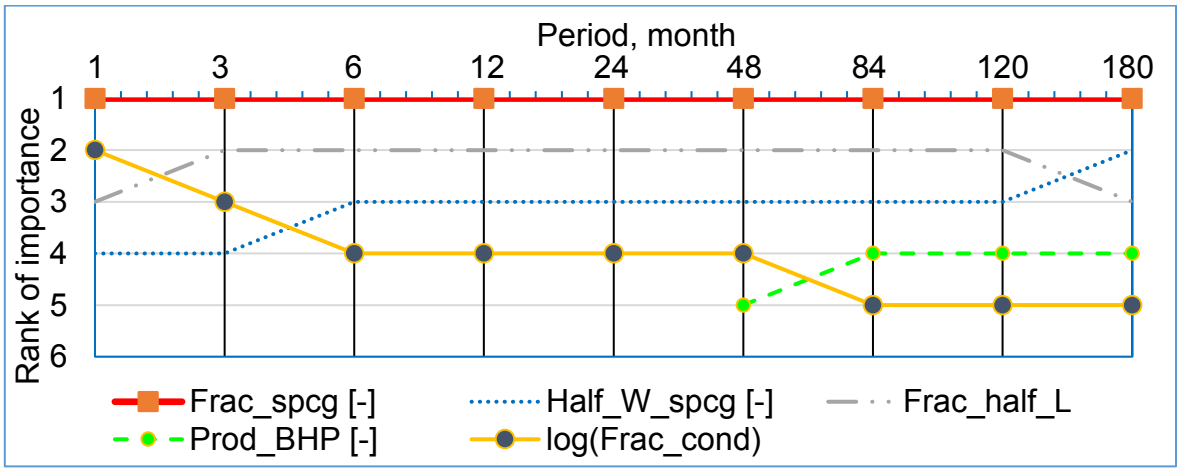


Fig. 3.5—Tracking the rank of important variables during the 15-year production period for PP.

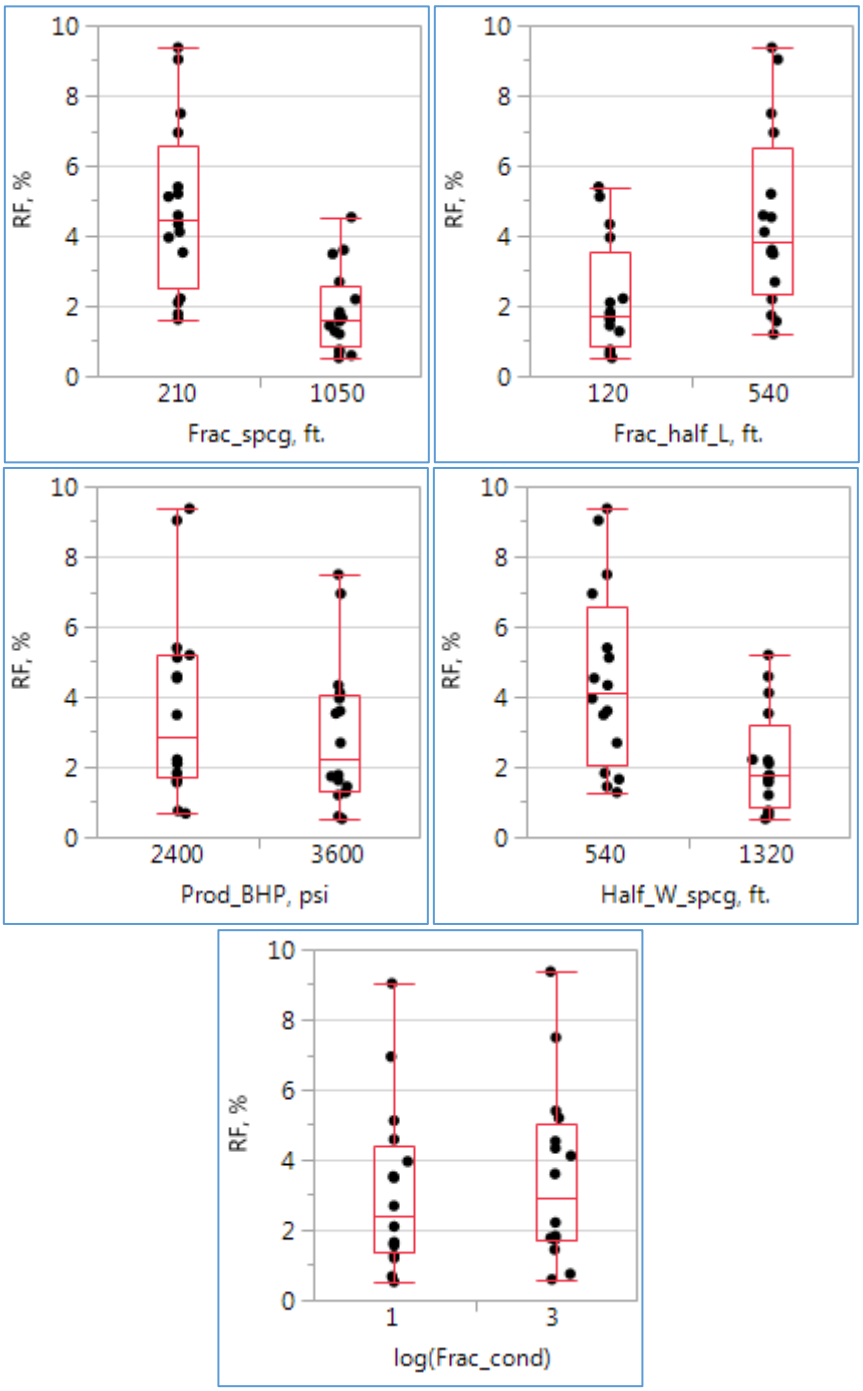


Fig. 3.6—Box-plots of the 15-year RFs of the 32 screening runs at both levels of the important variables of PP.

For all recovery schemes, fracture half-lengths and conductivities have positive main effects on RF, while fracture spacing, well spacing, and producing bottom-hole pressure have negative main effects. This means that RF increases when fracture spacing, well spacing, and producing bottom-hole pressure decrease, and when fracture half-lengths and conductivities increase within the investigated ranges. An important implication is that half-well spacing should be minimized (the minimum being equal to fracture half-length, which should be maximized). The author believes that there is no significant increase in RF when half-well spacing exceeds fracture half-length because the reservoir volume beyond the stimulated reservoir volume (SRV) is less affected by the depletion process for PP. In addition, it would be less accessible to the injected fluids for the three injection schemes because it is not along the path of steepest pressure gradient. This is especially true once breakthrough occurs. Thus for best performance, fracture half-length should be maximized and well spacing should be designed to be equal to double of fracture half-length. This, in addition to minimizing fracture spacing, accelerates the oil recovery.

Another key finding is that, for all recovery schemes, fracture conductivity is important (which agrees with Ye 2016) despite the common belief that fracture conductivity is not important for shale reservoirs. It is well known that dimensionless fracture conductivities (C_{fD}) above around 30 will cause the fracture to apparently have an infinite effective conductivity (Cinco-Ley and Samaniego 1981). Thus it would be uneconomical to increase the fracture conductivity to the point that C_{fD} exceeds 30. Based on that, and with the model's properties and assuming a fracture half-length of 540 ft, fracture permeabilities above 81 md (modeled fracture conductivity > 8.1 md.ft or

log(Frac_cond) > 0.91) are not required. However, this work shows that increasing fracture conductivity beyond that limit has a positive effect on RF. This could be attributed to the deeper induction of secondary fractures and microfractures beyond the main hydraulic fractures associated with larger conductivities. Such secondary fractures are expected to enhance and accelerate the flow of oil in shale. The two highest production rates in the first year have fracture conductivity of 1000 md.ft.

For injection schemes, injection time changes its effect from positive to negative later during production. In the beginning, starting injection decelerates production and RF becomes lower than their corresponding PP cases. However, with time, the effect of injection on production starts to appear and RFs of the injections schemes overtake those of PP. Overall earlier injection was found to have a better effect on accelerating recovery. Thus the focus in the optimization stage will be on earlier start of injection. This is also supported by **Fig. 3.7a**, which shows that the low level of injection time (one year) is preferred for all post-primary recovery schemes. For injection BHP, it has a positive effect on the 15-year RF for WF and CON, but its effect is not important on HNP. Fig. 3.7b shows that the higher level of injection BHP is preferred, which is expected.

For the two EOR cases (HNP and CON), minimum miscibility pressure (MMP) has a negative main effect on HNP and no significant effect on CON. The degree of miscibility (w) has no significant effect on either HNP or CON. The type of injection gas is less influential than the other important variables. Thus less miscible gases could be used and provide good results compared to the corrosive and expensive CO_2. However, both MMP

and w are present in some important interactions. Thus based on the effect heredity principle (Montgomery 2012), they will not be discarded in the optimization stage.

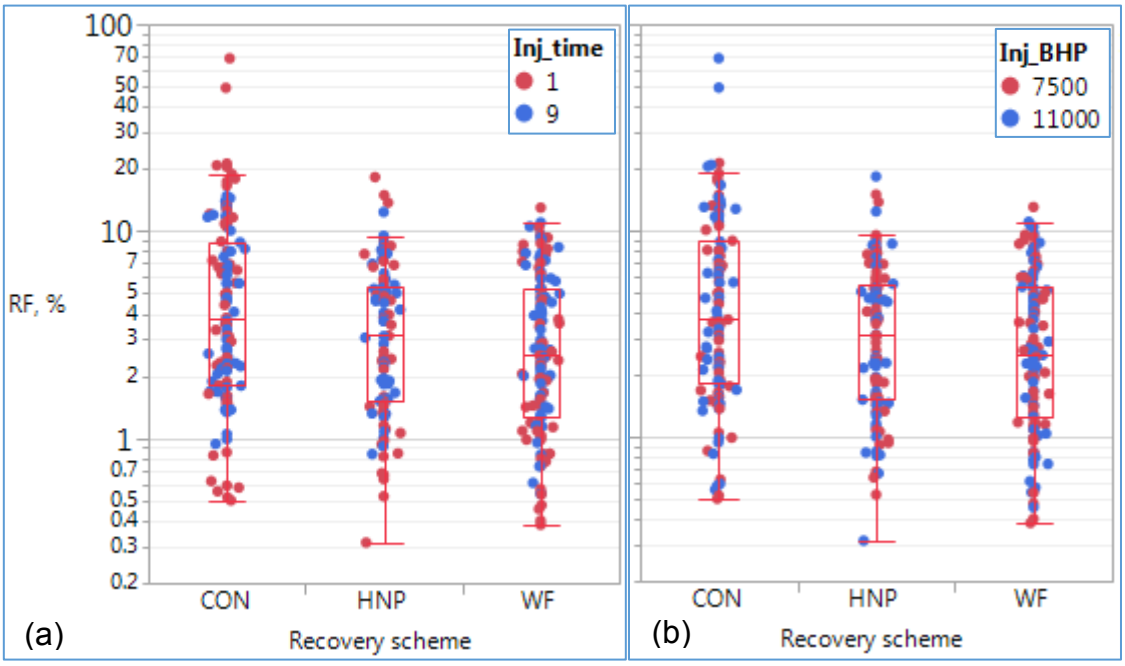


Fig. 3.7—Distribution of RF for the three injection schemes plotted by (a) Inj_time; (b) Inj_BHP.

For HNP, increasing the production duration (puff period) and reducing the injection period (huff period) increase the RF. This agrees with Chen et al. (2014). However, it is not believed that this conclusion is definitive because of the nature of fractional factorial design used (confounding exists). Further investigations are required if HNP performance is to be optimized. However, this is beyond the scope of this study. As stated previously, the optimization stage will not consider HNP.

3.3.2 Optimization

3.3.2.1 Primary Production (PP)

Optimization starts with redefining the ranges of the important variables selected based on the screening DoE results. For PP, the five investigated variables are important (Fig. 3.4). Therefore, the levels in the optimization stage will be modified accordingly to focus more on the levels of a variables that give higher RF. Two more levels are introduced to provide a better coverage of the variable space. The variables and their new ranges are summarized in **Table 3.5**.

Variable	Unit	Levels			
		1 (low)	2	3	4 (high)
Fracture spacing (Frac_spcg)	ft	90	360	630	900
Fracture half length (Frac_half_L)	ft	180	300	420	540
Half-well spacing (Half_W_spcg)	ft	540	720	900	1,080
Log of fracture conductivity (log(Frac_cond))		0.699 (5 md.ft)	1.466 (29.24 md.ft)	2.233 (171 md.ft)	3 (1,000 md.ft)
Producing BHP (Prod_BHP)	psi	2,400	2,700	3,000	3,300

Table 3.5—Modified ranges and levels of PP important variables for optimization.

The next step is to determine the adequate number of runs required for the optimization. However, this "adequate" number depends on the complexity of the relationship between the response and its factors. And since this relationship is not exactly known, this number is difficult to know a priori and will remain a mystery (Wang and Shan 2006; Johnson et al. 2011). That is why this study uses sequential DoE. The first stage of sequential DoE uses I-optimal design. The number of runs chosen is 100 because

it provides the maximum G-efficiency (**Fig. 3.8**), which is 0.68 for this case. This value lies within the recommended range (0.6 – 0.7) provided by Eriksson (2008). Furthermore, the number of runs per variable (20) is acceptable based on the literature findings (Chapman et al. 1994; Jones et al. 1998; Loeppky et al. 2009).

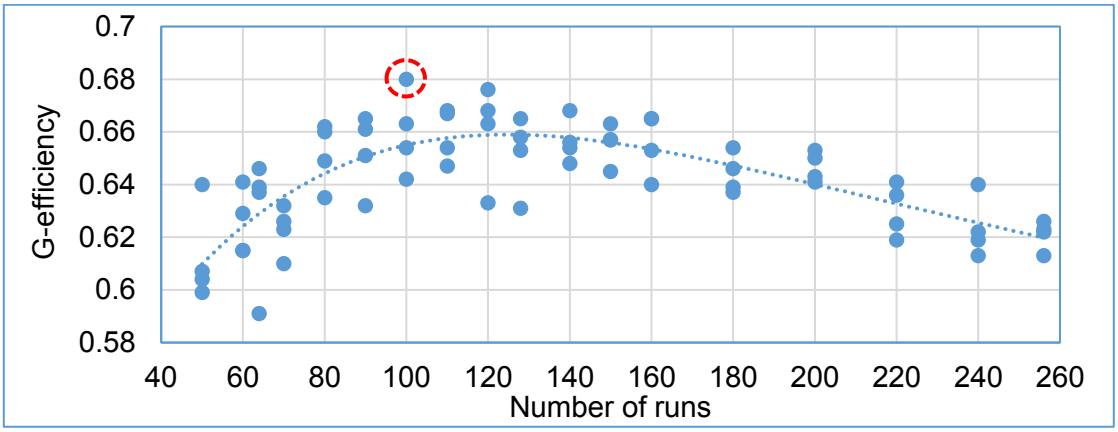


Fig. 3.8—G-efficiency variation with the number of runs generated by I-optimal design for PP.

Fig. 3.9 shows the distribution of RF after 15 years. The slopes of the blue lines joining the means supports the previous conclusions of important variables in the screening study. For example, the line in the fracture spacing boxplot has the maximum slope, which indicates that fracture spacing is the most important variable and that RF increases with its decrease. Such conclusions are also emphasized by the contour plots given in **Fig. 3.10**. In addition, not being parallel to the axes, the contour lines in Fig. 3.10 indicate the presence of interactions between the important variables, which agrees with the screening study conclusions. The contour plots also corroborate the previous finding that fracture half-

length should be maximized and well spacing should be designed to be equal to double of fracture half-length.

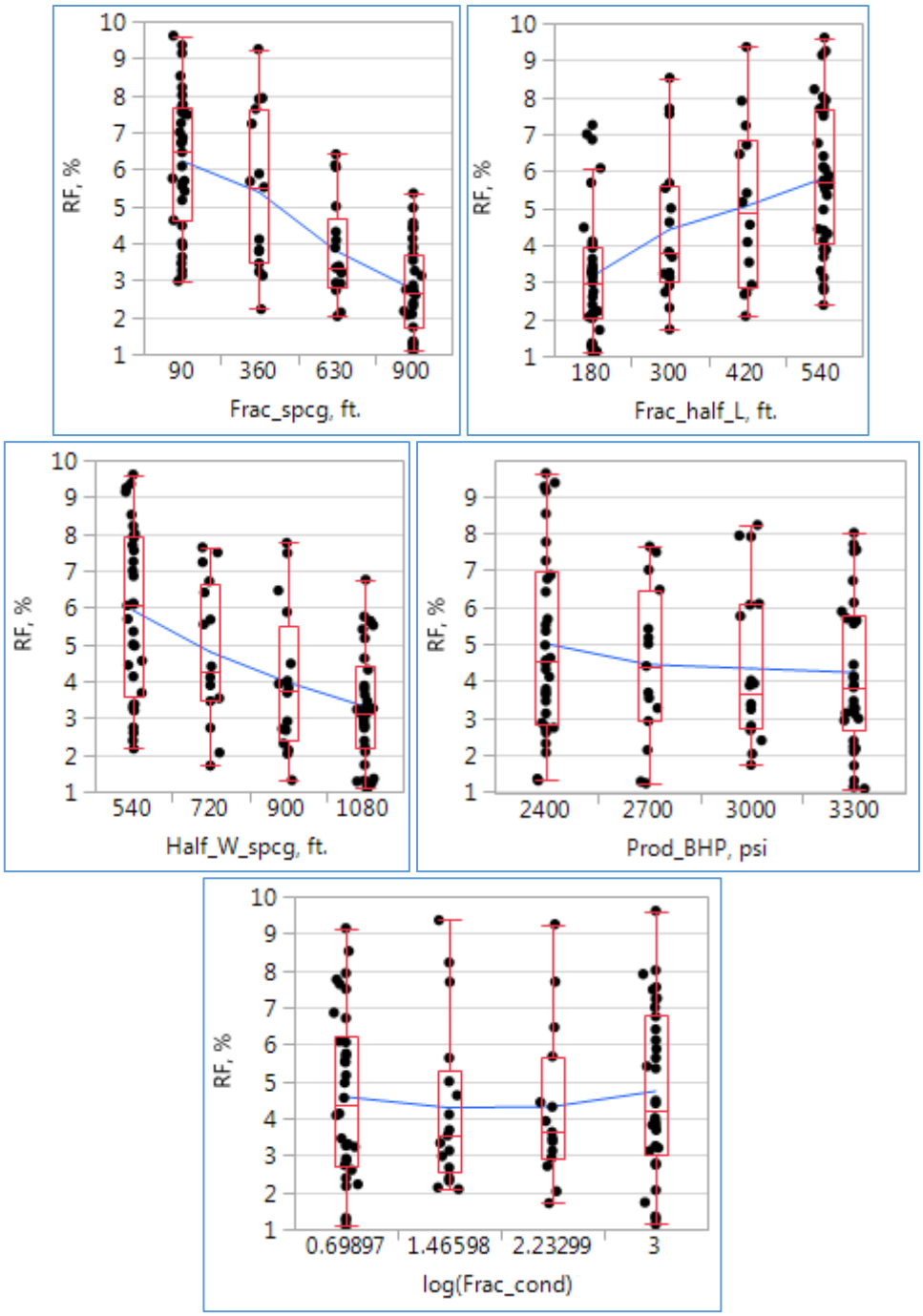


Fig. 3.9—Box-plots of the 15-year RFs of the 100 optimization runs at all levels of the important variables of PP (solid blue lines connect mean RFs).

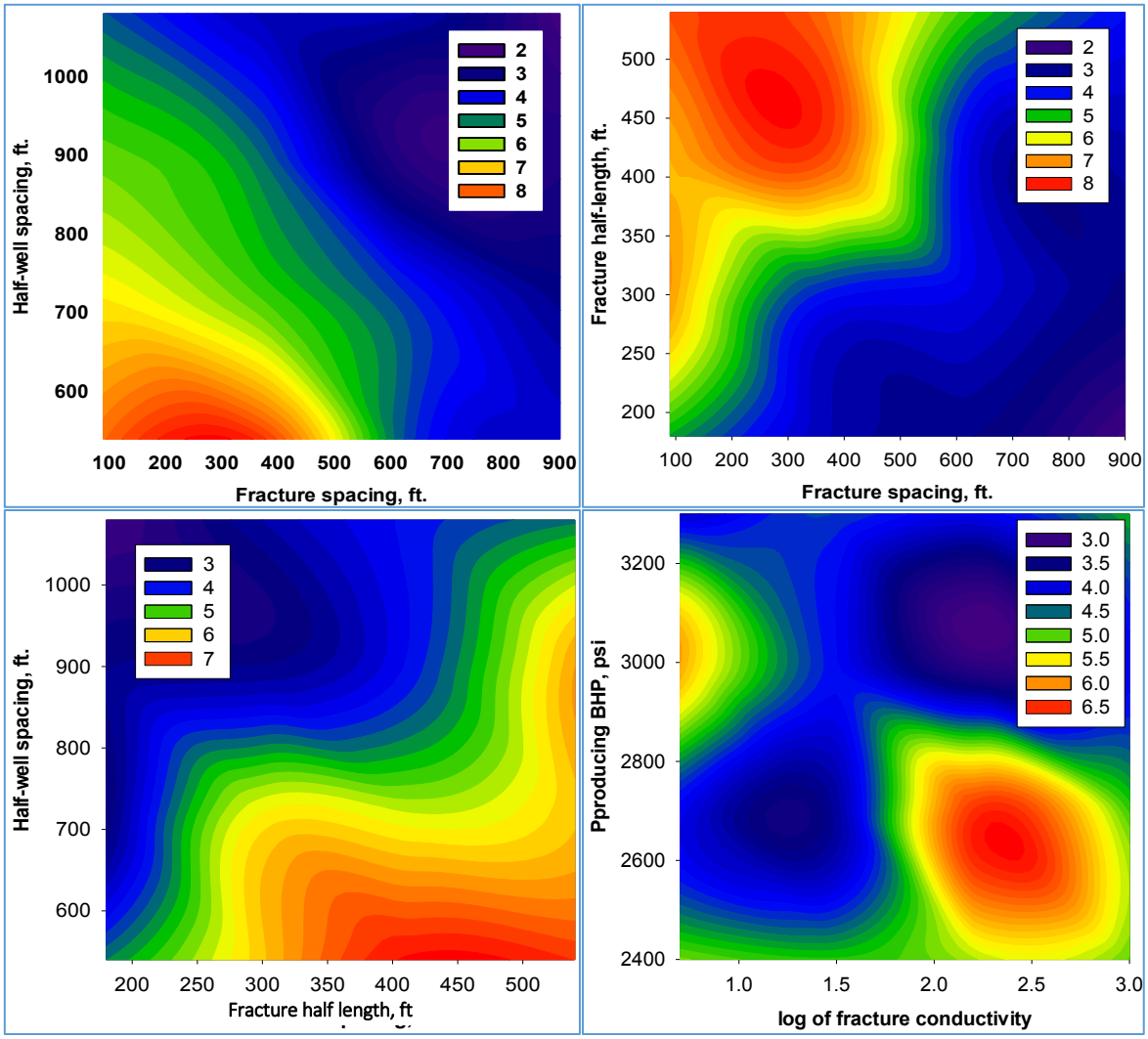


Fig. 3.10—Contour plots of the 15-year RF vs. the five important variables of PP.

Fig. 3.11 shows the RF change during the 15 years of production. The maximum RF is 9.6%. This maximum recovery case also has the maximum rate of RF increase. Its design settings are 90 ft for fracture spacing, 540 ft fracture half-length, 540 ft for half well-spacing, 2,400 psi for producing bottom-hole pressure, and 1,000 md.ft for fracture conductivity. The maximum RF conditions are constrained by the variable space specified

earlier (Table 3.5). The optimum conditions could be outside that design space. This is apparent in Fig. 3.9. Thus the second round of sequential DoE will extend the search for the optimum outside the design space specified in Table 3.5. To facilitate the search for the optimum however, the following points are important to consider:

1. As shown earlier, fracture conductivity is an important variable and having conductivities of 1,000 md.ft (or even more) accelerates the recovery. As a result, fracture conductivity is kept at its maximum of 1,000 md. This value is not exceeded because it is hard to go beyond it from the practical viewpoint.
2. The proportion of reservoir volume affected by main hydraulic fractures and their induced fractures will be larger for closely spaced fractures. Thus fracture spacing range is extended below 90 ft to search for the optimum.
3. Producing BHP is restricted to be above the bubble point pressure of 2,375 psi (this might need to be revised though). Therefore, it is kept at its minimum, 2,400 psi.
4. Fracture half-length will be extended beyond its current maximum limit of 540 ft. Based on the previous findings, half-well spacing will be set equal to fracture half-length. Thus half-well spacing is not included as an independent variable.
5. It is assumed that the effects of interactions will be the same.

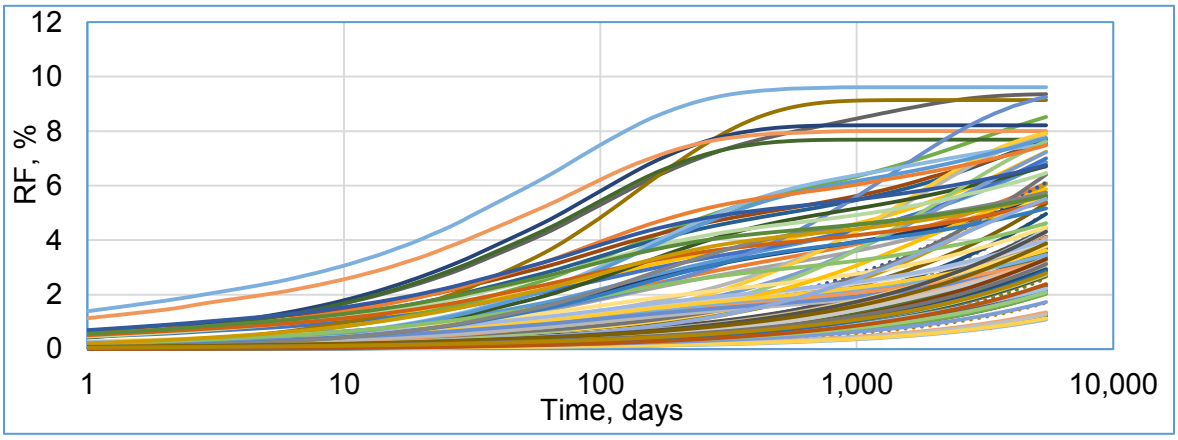


Fig. 3.11—Change of RF of the 100 optimization PP runs with time.

With only fracture half-length and fracture spacing selected for the second optimization round of sequential DoE, five levels are chosen for each of them (**Table 3.6**). A full factorial design ($5^2 = 25$ runs) is conducted. The 3D and contour plots of the 15-year RF are provided in **Fig. 3.12**. It shows that for maximum RF, fracture spacing should lie between 20 and 40 ft. In addition, the contour lines are almost parallel to the fracture half-length axis. This implies that 15-year RF is not sensitive to fracture half-length or its interaction with fracture spacing in this region of the variable space. This could be attributed to the fact that half-well spacing is set equal to fracture half-length. If larger half-well spacing is used, RF would depend on both half-well spacing and fracture half-length (as was the case in the screening study). However, despite being uninfluential on the RF, increasing fracture half-length reduces the number of wells that need to be drilled. Thus the economics of shale oil production shall be improved.

Variable	Unit	Levels				
		1	2	3	4	5
Fracture spacing (Frac_spcg)	ft	90	70	50	30	10
Fracture half length (Frac_half_L)	ft	540	660	780	900	1,020

Table 3.6—Modified ranges and levels for second round of PP optimization in sequential DoE.

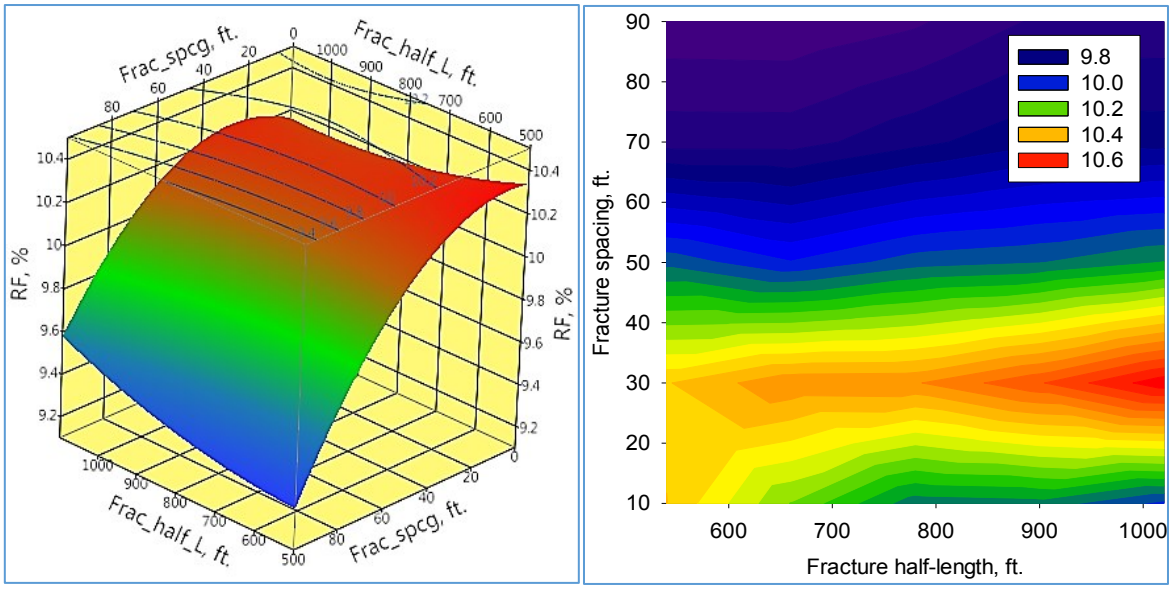


Fig. 3.12—3D and contour plots of the 15-year RF for the second round of PP optimization.

The maximum RF among the 25 runs is 10.6%. It corresponds to 30 ft fracture spacing. This implies that PP optimization is not enough to improve the RF from Eagle Ford shale. It can only accelerate the recovery and cash flow. Its effect is limited on improving the primary RF, which has a maximum physical limit (determined by pressure depletion) that cannot be exceeded by PP optimization alone. Exploring other recovery methods is a necessity.

3.3.2.2 Waterflooding (WF)

Like PP, WF optimization starts with redefining the ranges of variables. Based on our understanding of the whole recovery process, the problem is simplified by reducing the number of variables to be optimized as follows. It is believed that keeping producing BHP at its minimum (2,400 psi) and fractures (injection and producing) conductivity at their maximum (1,000 md.ft) will enhance the production performance. To further simplify the problem, the lengths of the injecting and producing fractures are treated as one variable. Like PP, half-well spacing is set equal to fracture half-length. In addition, injection BHP is kept at its maximum (11,000 psi), which will maximize RF (its important interactions with other variables have synergistic effect). Therefore, only three variables (fracture spacing, fracture half-length, and injection time) are left to be optimized instead of the important nine defined by the screening study. Their levels are given in **Table 3.7**. The levels of fracture spacing and fracture half-length are chosen based on the knowledge and understanding developed thus far from PP. This was also verified by applying one-factor at a time (OFAT), which showed that the maximum RF lies in the same range as that of PP for fracture half-length and fracture spacing. For injection time, the study focuses on the range of 0.0 to 2.0 years because of the rapid decline in PP production, which requires pressure support to start as soon as possible.

A full factorial design ($5^3 = 125$ runs) is done and 3D plots of the 15-year RF are provided in **Fig. 3.13**. Like PP, the contour lines are parallel to the fracture half-length axis and therefore, RF does not depend on fracture half-length or its interaction with fracture spacing. This supports the previous justification that RF is not sensitive to fracture half-

length as long as half-well spacing is set equal to fracture half-length. The same applies to injection time. This could be because most of the recovery occurs before the end of the 15-year production period, which makes RF at 15 years relatively insensitive to injection time. However, increasing the fracture half-length and starting the injection early would improve the economics of shale oil production because of the earlier cash flow associated with higher early production. For fracture spacing, its continuous decrease from 90 to 10 ft improves RF. The optimum PP fracture spacing of 20 to 40 ft is economically reasonable and agrees with the current practical estimates. These observations imply that the optimum design of lateral spacing and hydraulic fractures is the same for PP and WF. The maximum RF for WF is nearly 54.4%, almost five times as much as that of PP (10.6%). This number is unrealistic and could be due to the assumption of homogeneous reservoir without any natural fractures. The results would not be as promising if natural fractures are included. The systematic DoE workflow can be further extended to explore different models.

Variable	Unit	Levels				
		1	2	3	4	5
Fracture spacing (Frac_spcg)	ft	90	70	50	30	10
Fracture half length (Frac_half_L)	ft	540	660	780	900	1,020
Production time after which injection starts (Inj_time)	years	0	0.5	1	1.5	2

Table 3.7—Modified ranges and levels for WF optimization.

3.3.2.3 Continuous Gas Injection (CON)

Based on the previous findings and conclusions, the interest here is restricted to four variables: fracture spacing, injection time, degree of miscibility, and minimum

miscibility pressure. Their levels are given in **Table 3.8**. Half-well spacing is set equal to fracture half-length. A full factorial design (4*4*3*3 = 144 runs) is done.

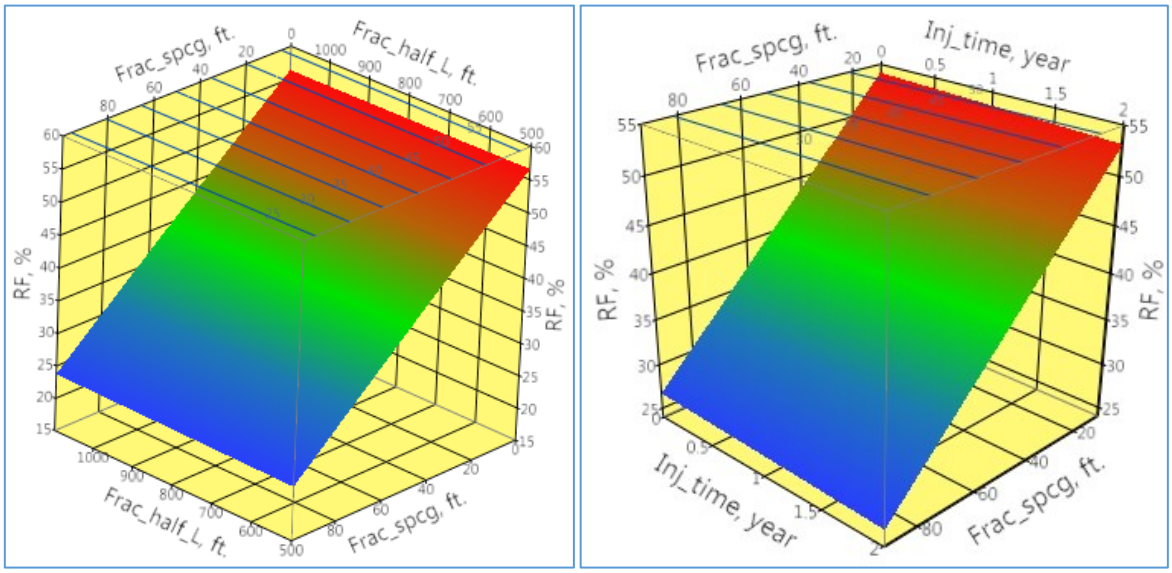


Fig. 3.13—3D plots of the 15-year RF for WF optimization.

Variable	Unit	Levels			
		1	2	3	4
Frac spacing (Frac_spcg)	ft	10	30	50	70
Production time after which injection starts (Inj_time)	Year	0	0.5	1	1.5
Degree of miscibility (w)		0.2	0.5		0.8
Minimum miscibility pressure (MMP)	psi	2,500	3,750		5,000

Table 3.8—Modified ranges and levels for CON optimization.

The 15-year RF of all 144 runs is almost 100%. This could mean that RF is not sensitive to any of the four variables considered because all runs have the same RF regardless of the variable settings. However, earlier injection, higher degree of miscibility, and lower MMP are preferred from the economic viewpoint because they attain the same RF earlier. Likewise fracture spacing between 20 and 40 ft is better. This is an important

finding because the optimum design (20 – 40 ft fracture spacing and well spacing that is twice as much as fracture half-length) is the same for all recovery schemes. The initial optimum design for PP will still be optimum for subsequent IOR injection schemes. No infill wells or refracturing is required.

Certainly 100% RF is not realistic. However, having half-well spacing equal to fracture half-length maximizes the volume swept by injection gas because all the simulated volume lies within the path of steepest pressure gradient between the injecting and producing fractures along the same lateral. Thus gas sweeps all the SRV. This explains why miscible gas flooding has the potential of achieving RF close to 100%. However, the presence of heterogeneity and natural fractures would slow down the recovery process and cause early breakthrough. RF would be much less than these unrealistically high values. This needs further investigation. Actually, Sheng and Chen (2014) had a comparable model structure and obtained 99.3% RF. Other studies (e.g., Zhu et al. 2017) with similar model structure and the same assumptions (homogeneous reservoir with no natural fractures) did not achieve 100% RF. This is probably because they did not use a well spacing that was twice as long as fracture half-length. It is important to recognize that the presented completion design might not be logistically feasible and it is not clear yet how to implement it. Drilling two laterals alongside each other might be more feasible but this needs more logistic studies.

3.4 Conclusions

The main goal of this chapter was to illustrate how to use simulation and design of experiments (DoE) to maximize RF of four recovery schemes (primary production (PP),

waterflooding (WF), continuous miscible gas injection (CON), and huff and puff miscible gas injection (HNP) in a homogeneous Eagle Ford reservoir with no natural fractures. The objectives are to determine their most important variables, optimize their design, and compare their performance to specify the most prospective one. Adopting the innovative design of alternating injecting and producing fractures along the same lateral, the main findings and conclusions of the study are as follows:

1. PP optimization can only accelerate the recovery and cash flow from Eagle Ford shale. Its effect is limited on improving the primary RF, which has a maximum physical limit that cannot be exceeded by PP optimization alone. Thus exploring other recovery methods is a necessity to improve RF.
2. Continuous miscible gas injection has the highest potential to improve RF.
3. Developing alternating injection/production fractures along the same lateral might be the next breakthrough to boost the RF from shale plays.
4. Being two of the most important design variables for all recovery schemes, fracture spacing and fracture half-length should be better estimated in order to reduce the uncertainty in forecasting oil production.
5. For all schemes, the optimum design settings are as follows:
    a) Fracture spacing: should be from 20 to 40 ft.
    b) Fracture half-length: should be maximized as practically as possible.
    c) Well-spacing: should be twice as long as fracture half-length. In such a way, all the reservoir volume between wells will be covered with hydraulic fractures. This does not consider the fracture hit effect.

d) Fracture conductivity: is important for shale oil recovery despite the common belief that it is not important. It should be maximized.

e) Bottom-hole pressure: should be minimized (above the bubble point pressure).

6. For secondary and tertiary recoveries to achieve the best performance, injection should start as early as possible (preferably, when the well is put on production) with the maximum injection pressure.
7. For CON and HNP, although minimum higher miscibility pressure (MMP) and degree of miscibility are preferred in general, the type of injection gas is not as influential as the fracture design and well spacing. Less miscible gases could be used and provide good results.

4. ECONOMIC OPTIMIZATION OF PRIMARY OIL RECOVERY IN BAKKEN SHALE USING STATISTICAL DESIGN OF EXPERIMENTS

4.1 Introduction

Despite leveraging the sophisticated technology of horizontal wells and multistage hydraulic fracturing, recovery factor (RF) is still very low. In the US, the RF in the producing tight oil fields can be as low as 1% and as high as 12% with the normal range being from 3 to 7% of the original oil in place (Joshi 2014). In addition, the high capital expenses of the massive stimulation treatments and horizontal drilling add to economic stress. Furthermore, the high uncertainty in many controllable completion design variables and uncontrollable reservoir variables exacerbates the economic uncertainty. Thus economic optimization of oil recovery and hydraulic fracturing treatments is of utmost importance. Again DoE is used here to maximize the economic oil recovery and quantify the uncertainties through efficiently and systematically exploring the multidimensional variable space.

In Chapter 3, an optimization DoE-based workflow was used to optimize well and fracture design for maximum primary and enhanced oil recovery in Eagle Ford shale. The effects of uncontrollable reservoir variables (e.g., permeability) were not considered. Thus this chapter considers these variables in addition to the controllable design variables to compare their relative importance for shale oil recovery. In addition, economics is considered here with net present value (NPV) as the response to be maximized instead of

RF. Finally, reservoir heterogeneity is considered to provide a more realistic reservoir description.

4.2 Methods

Fig. 4.1 summarizes the main steps of the workflow. Net present value (NPV) is selected as the response to be optimized. A thorough literature review identifies 16 uncertain variables which are varied in a screening DoE study to determine which ones are more significant for NPV. These significant "heavy hitters" will be incorporated into an optimization DoE to obtain a response surface (RS). This semi-empirical RS is then used to optimize the economic recovery by maximizing NPV, and to evaluate the uncertainty in the NPV using Monte Carlo (MC) simulation. JMP and R statistical software are used here for statistical design and analysis.

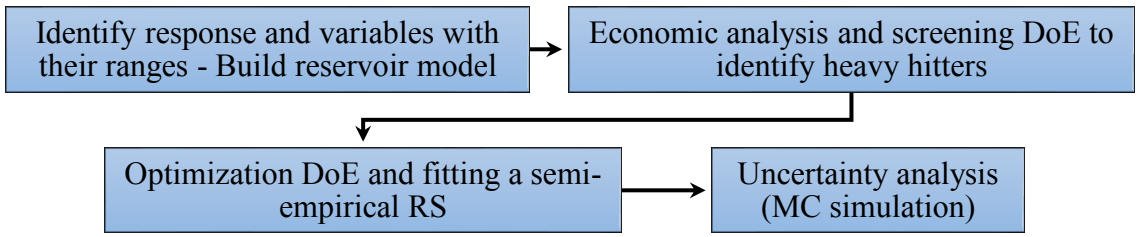


Fig. 4.1—Flowchart of economic optimization approach.

4.2.1 Simulation Model Description

The general structure of the model is the same as Eagle Ford shale model (section 3.2.2 and Fig. 3.2). CMG is used for modeling. A thorough literature review (Shoaib and Hoffman 2009; Nojabaei et al. 2012; Cherian et al. 2013; Dong and Hoffman 2013; Kurz

et al. 2013; Morsy et al. 2013; Wan et al. 2013a, 2013b; Xu and Hoffman 2013; Chen et al. 2014; Eshkalak et al. 2014; Fai-Yengo et al. 2014; Gamadi et al. 2014; Han and Gu 2014; Liu et al. 2014; Saputelli et al. 2014; Sheng and Chen 2014; Wan et al. 2014; Yu and Sepehrnoori 2014b; Yu et al. 2014; Alharthy 2015; Fragoso et al. 2015; Kalra and Wu 2015; Pu and Li 2015; Rivera et al. 2015; Wan et al. 2015; Zhu et al. 2015) has been done to identify 16 uncertain variables for Middle Bakken shale, five controllable completion design variables and 11 uncontrollable reservoir variables (**Table 4.1**). Factor ranges are chosen carefully to avoid the dominance of numerical errors on the response and the risk of overlooking a meaningful factor. Due to their wide ranges, matrix permeability, vertical to horizontal permeability ratio, and fracture conductivity are expressed logarithmically.

The effect of heterogeneity is incorporated into the model through Dykstra Parsons coefficient (DP). Permeability (k) is assumed to have log-normal distribution. In other words, $Y = \ln(k)$ has normal distribution with mean μ and standard deviation σ, which are related to heterogeneity, mean permeability (E(k)), and permeability variance (Var(k)) through the following equations (Zhu et al. 2015):

$$DP = 1 - \exp(-\sigma) \quad \text{(Eq. 4.1)}$$

$$E(k) = \exp(\mu + \sigma^2/2) \quad \text{(Eq. 4.2)}$$

$$Var(k) = (\exp(\sigma^2) - 1) \times \exp(2\mu + \sigma^2) \quad \text{(Eq. 4.3)}$$

Unlike the deterministic cases in Chapter 3, using inferential statistics is more justifiable here because introducing heterogeneity ensures randomness in the simulation model.

Type	#	Variable	Unit	Levels	
				Low	High
Controllable Design Variables	1	Half well spacing (Half_W_spcg)	ft	540	1,320
	2	Prod. frac half length (Frac_half_L)	ft	120	540
	3	Frac spacing (Frac_spcg)	ft	210	1,050
	4	Frac conductivity (log(Frac_cond))	md.ft	1 (0)	100 (2)
	5	Producing BHP (Prod_BHP)	psi	1,600	3,000
Uncontrollable State Variables	6†	Initial reservoir pressure (P_res) †	psia	5,000	7,000
		Reservoir temperature	deg F	225	255
		Depth	ft	9,500	11,500
	7	Matrix porosity (Poro)	%	2	10
	8	Matrix permeability (log(Perm))	md	0.0005 (-3.30103)	0.05 (-1.30103)
	9	Dykstra-Parsons coeff. (DP)		0	0.8
	10	K_v/K_h (log(K_v/K_h))		0.01 (-2)	1 (00)
	11	Rock compressibility (R_compr)	$10^{-6}psi^{-1}$	3.5	8.5
	12	Pay thickness (h)	ft	30	70
	13	Initial connate water saturation (S_{wi})	ft	0.25	0.45
	14	Residual oil saturation (S_{or})		0.1	0.4
	15	Matrix endpoint relative permeability to water at S_{or} (K_{rw})		0.024	0.6
	16	Matrix endpoint relative permeability to oil at S_{wi} (K_{row})		0.1	1

† Only initial reservoir pressure was considered to avoid multicollinearity problems. Reservoir temperature and depth were modified accordingly.

Table 4.1—Screening controllable and uncontrollable variables and their ranges.

For the oil model, Bakken oil composition (**Fig. 4.2**) is taken from Pu (2013). Peng-Robinson Equation of State (EOS) is employed to calculate fluid properties (**Fig. 4.3**) and phase envelope (**Fig. 4.4**).

For DoE and RSM, more details will be provided in section 4.3 since it is a sequential process (i.e., each round of DoE depends on the results of the previous rounds).

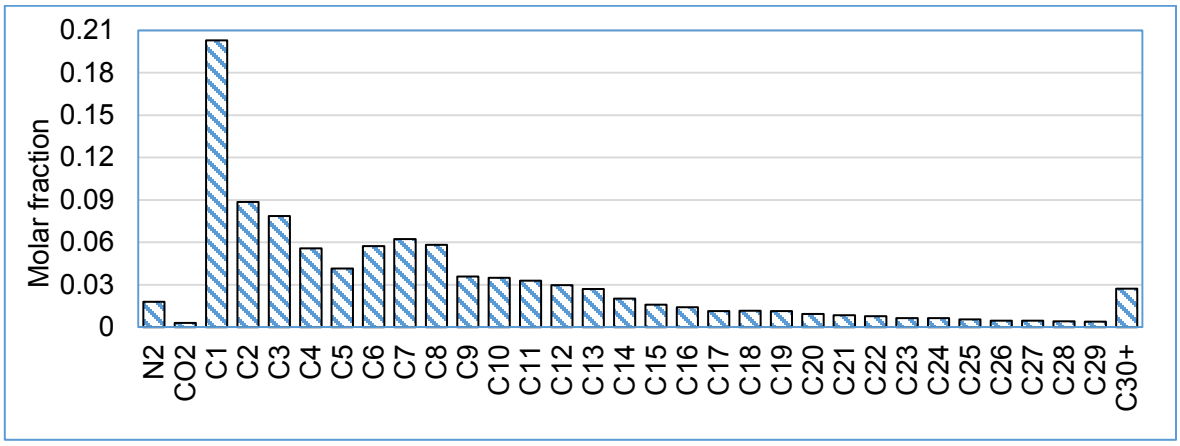


Fig. 4.2—Bakken reservoir oil composition.

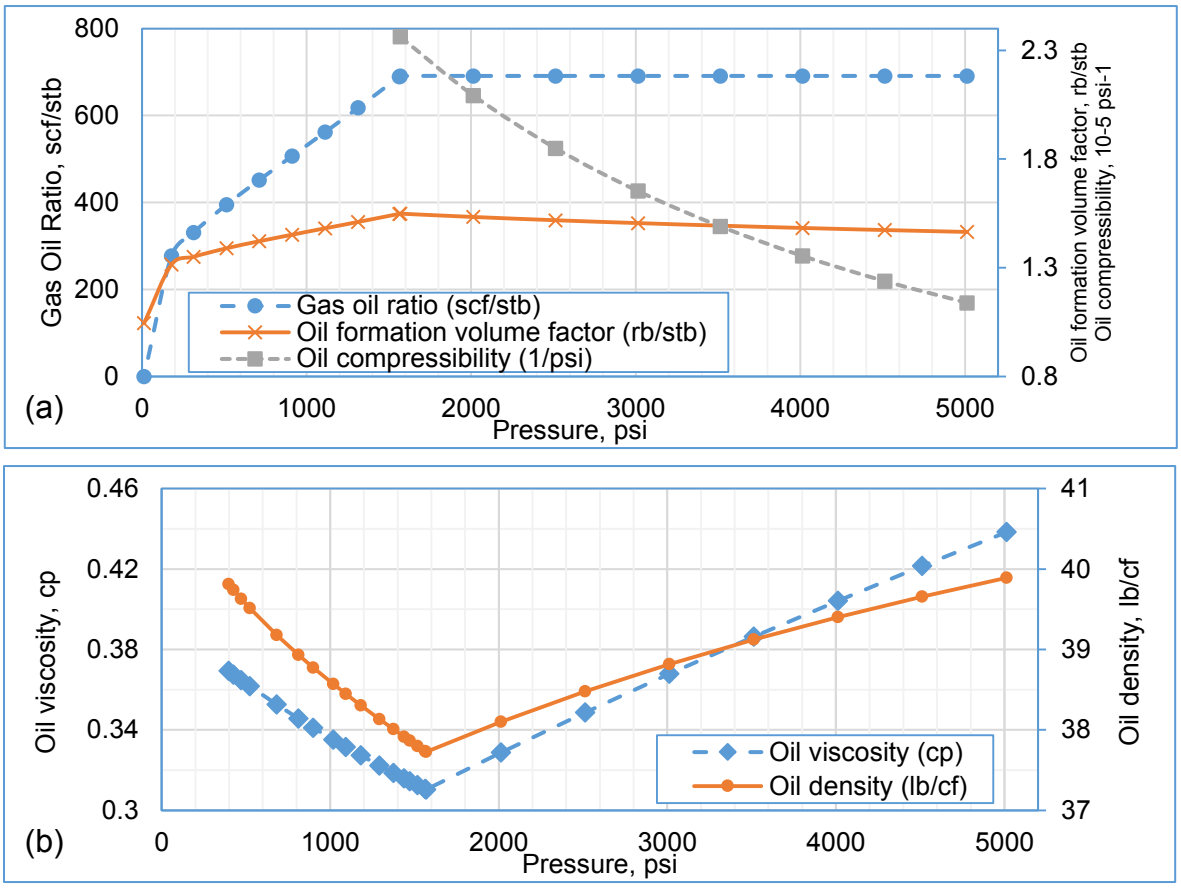


Fig. 4.3—Bakken reservoir oil properties. (a) Gas oil ratio, oil formation volume factor, and oil compressibility vs. pressure; (b) Oil density and viscosity vs. pressure.

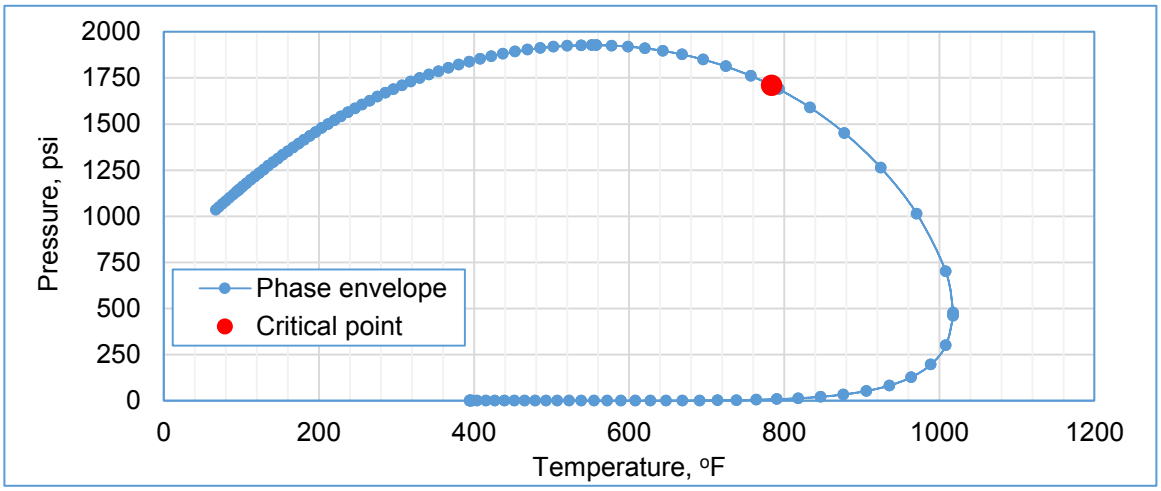


Fig. 4.4—Bakken reservoir oil phase envelope.

4.2.2 Economic Analysis

The NPV is calculated for 10 years based on a 1,280-acre producing area assuming a constant average oil and gas prices of USD 65/bbl and USD 3/Mscf, respectively. Annual interest rate, royalty taxes, and other taxes are assumed to be 10%, 12.5% and 11.5%, respectively (SPE 2015). Operating expenses (OPEX) are taken to be USD 20/bbl (EIA 2016c). Based on these assumptions, the net profit per barrel of oil equivalent (BOE) is USD 35.77. The cost of drilling a 10,080 ft lateral (a typical lateral length in Bakken shale (EIA 2016c)) is USD 4.80 MM. The cost of one fracture stage is calculated using the following equation (based on the work of Schweitzer and Bilgesu (2009)):

$$\text{Cost per stage (USD M)} = 75 + 0.1 \times \text{Frac\_half\_L} \quad \text{(Eq. 4.4)}$$

The calculated capital expenses per well agree with the literature (Schlumberger 2013; Apaydin 2014; Yu and Sepehrnoori 2014a; EIA 2016b, 2016c). The cumulative NPV is calculated based on the following equation:

$$\text{Cum NPV} = \sum_{\text{day (j)}=1}^{3653} \left(\frac{(\text{Oil and Gas prices} - \text{OPEX}) \times (1 - \text{Tax})}{(1 + \text{interest/day})^{j}} \right) - \text{CAPEX} \quad \text{(Eq. 4.5)}$$

where the daily interest rate is based on 10% annual interest rate.

4.3 Results

4.3.1 Screening (Sensitivity)

A sensitivity study is conducted to screen out the most significant variables (heavy hitters) from the pool of 16 variables selected previously. The response variables here are RF and NPV. For screening purposes, only two levels of the variables are of interest (Table 4.1). A full combination of all possible scenarios would necessitate 2^{16} = 65,536 simulation runs. However, this part of the study is only concerned with evaluating the main effects of the variables not their interactions. Two-level fractional factorial design can help reduce the number of runs considerably and provide the required information about main effects. This design is one of the most widely used in screening studies (Box et al. 2005; Mathews 2005; Eriksson 2008; Montgomery 2012; Jamshidnezhad 2015).

Resolution IV screening design (64 runs) is selected because it can estimate the main effects which will be aliased with three-factor interactions in the worst case. Assuming that three-level and higher interactions are negligible, all main effects are estimable with no confounding by using only 64 runs. This represents around 0.1% of the full factorial runs.

After running the 64 simulation scenarios and conducting the economic analysis, the minimum, median, average, and maximum NPV values are USD -44.99 MM, USD -11.33 MM, USD -10.45 MM, and USD 102.89 MM, respectively. Only 14 runs (21.88%) of the 64 runs yielded positive NPV. This signifies the stressing need to optimize shale oil

recovery process. It is remarkable to note that out of these 14 runs, all have the high permeability level (0.05 md) and 12 have the high porosity level (10%) indicating the importance of these variables.

NPV results are then used to construct a half normal probability plot (**Fig. 4.5**), which shows that there are eight significant variables along with some interactions affecting the 10-year NPV (those most deviated from the blue reference line). All of them (except the interaction term Half_W_spcg*Frac_spcg) have a positive effect on the NPV. These eight variables are directly related to reservolir oil in place (e.g., porosity and thickness) and/or productivity (e.g., permeability and fracture spacing)

Fig. 4.6 tracks how the ranks of the NPV significant variables and their interactions change with time based on their p-value. At early production time, half well spacing (Half_W_spcg) and fracture spacing (Frac_spcg) are the two most significant variables. Their ranks retreat back continuously, however, as production goes on, and they get replaced by porosity (Poro) and permeability (log(Perm)). Design variables initially dominate the effect on production economics and reservoir characteristics take over later (after one to two years) in the life of the well. **Fig. 4.7a** shows that NPV increases when the number of wells and fracture stages decrease (i.e., Half_W_spcg and Frac_spcg increase) within the tested range. Note that the presence of curved contour line agrees with the previous conclusion that interaction exists between these two variables represented with the term Half_W_spcg*Frac_spcg.

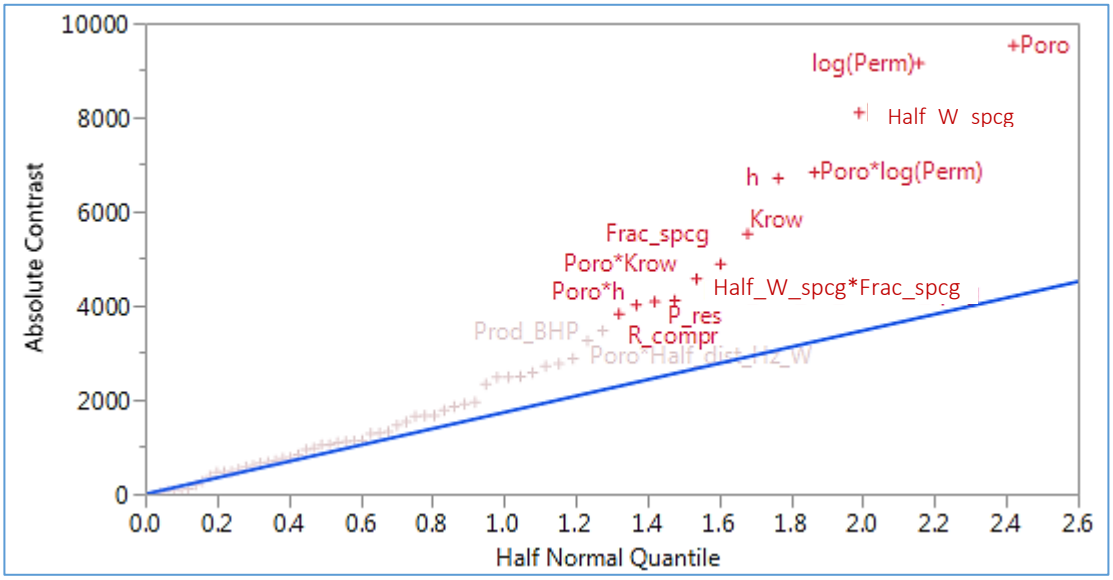


Fig. 4.5—Half-normal probability plot showing the most significant variables and interactions affecting the 10-year NPV.

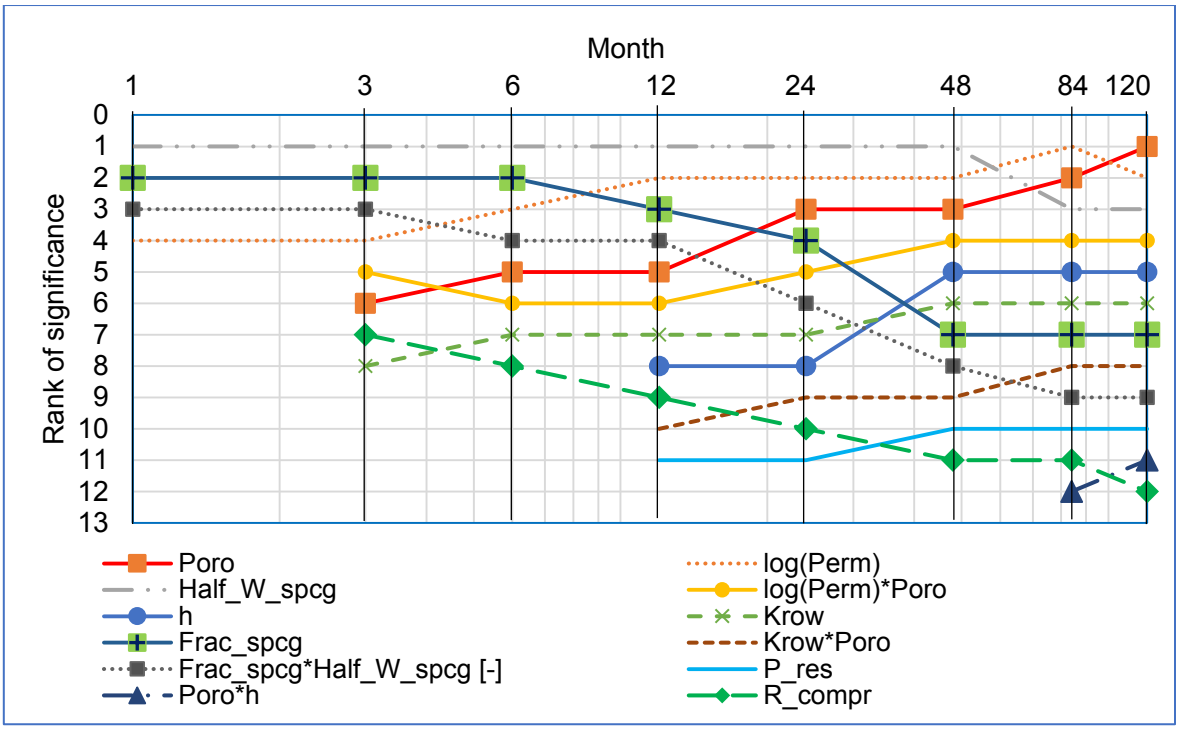


Fig. 4.6—Tracking the rank of NPV significant variables during production time. [-] means negative effect.

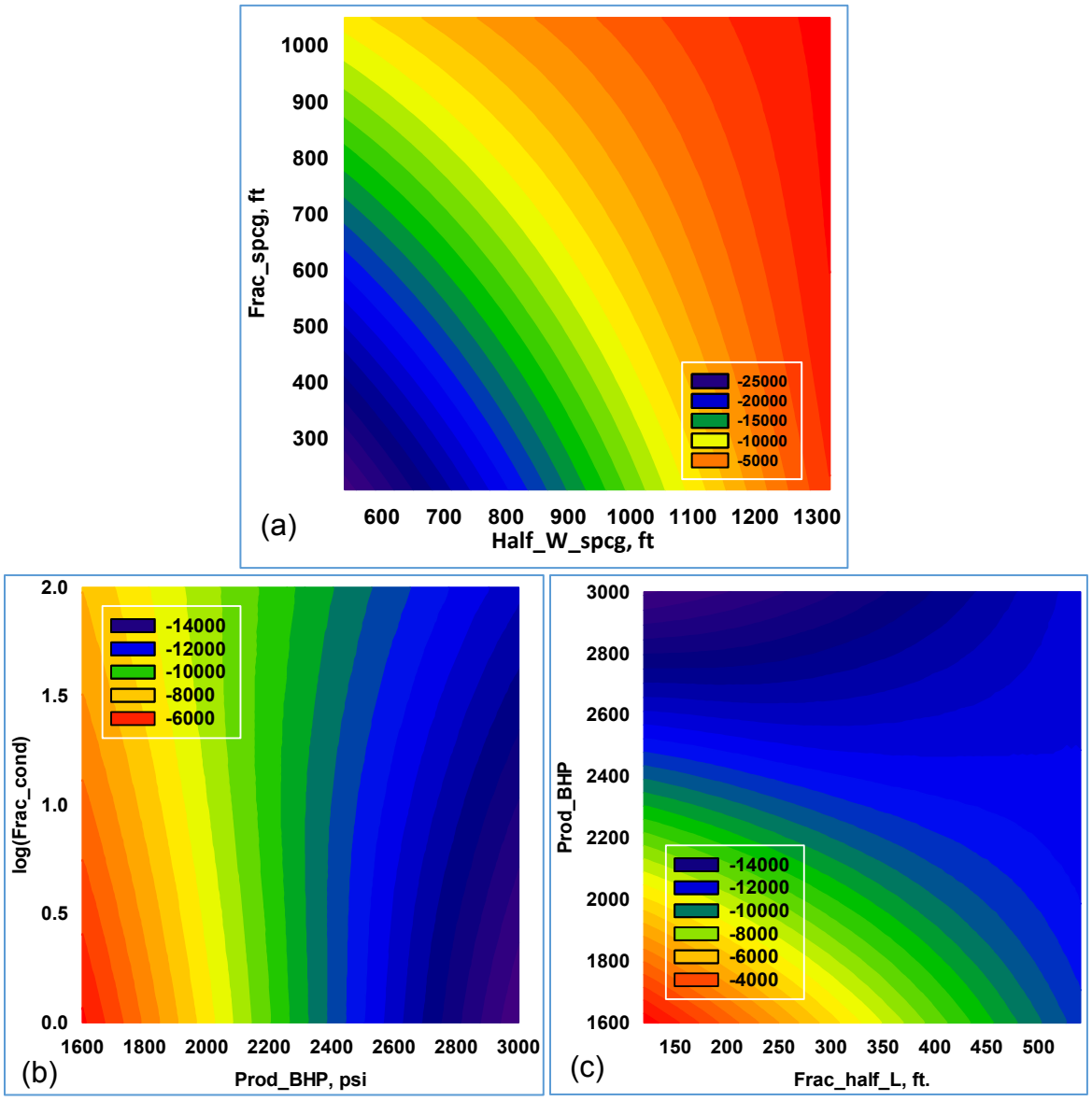


Fig. 4.7—Contour plot of 10-year NPV (in USD M) vs. (a) Half_W_spcg and Frac_spcg; (b) Prod_BHP and Frac_cond; (c) Prod_BHP and Frac_half_L.

It is noteworthy to observe that the less significant design variables (Prod_BHP, Frac_half_L, and log(Frac_cond)) affect the NPV. Their effect, however, is less pronounced than the effects of the significant variables. This should be considered while searching for the optimum conditions.

The same analysis of Fig. 4.6 could be done on the significant variables for RF after 10 years (**Fig. 4.8**). The uncontrollable reservoir variables and their interactions dominate the effect on RF throughout the life of the well. Because of such dominance, three of the five design variables (fracture half-length, farcture conductivity, and producing bottom-hole pressure) are not significant for NPV or RF.

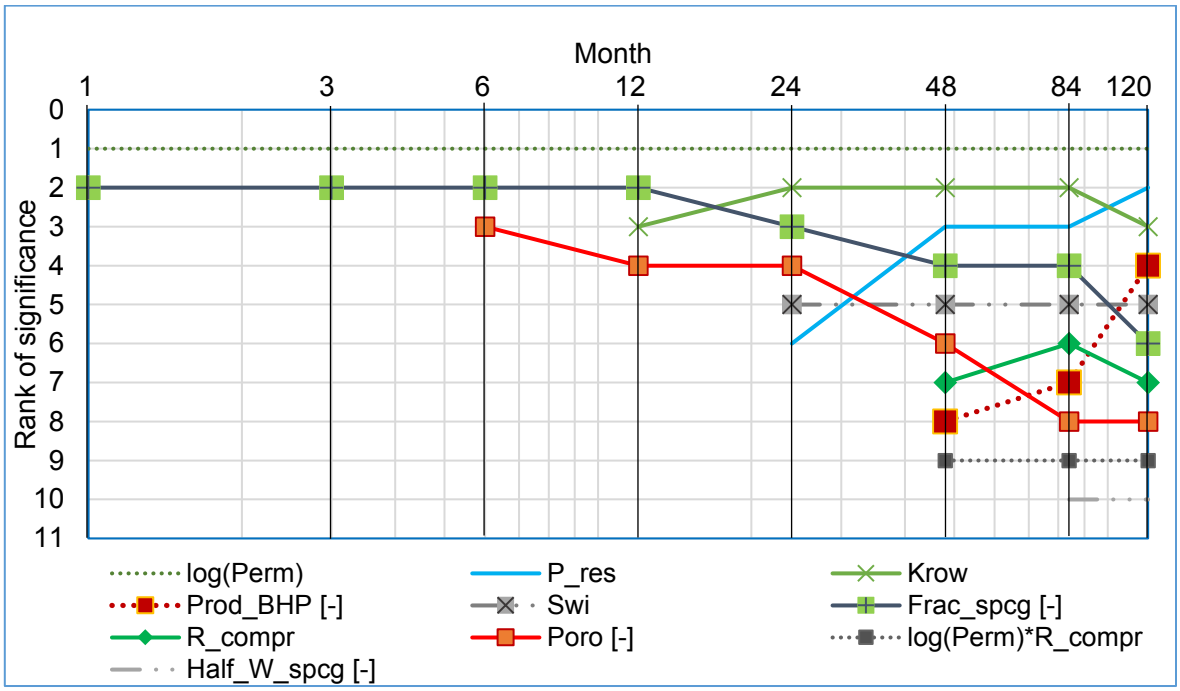


Fig. 4.8—Tracking the rank of significant variables during production time for RF. [-] means negative effect.

Statistically fracture half length (Frac_half_L) is not significant since its p-values are larger than 0.05. However, Fig. 4.7c demonstrates that reducing Frac_half_L would have a positive impact on the 10-year NPV. This could be because additional economic benefit from increasing oil production by increasing Frac_half_L does not cover the

additional capital expenses of the additional increase in Frac_half_L. The same is true for Prod_BHP (Fig. 4.7b).

Taking these observations into account in the optimization stage, the capital costs could be reduced and the Bakken oil recovery economics could be improved by:

- Reducing the number of drilled wells per unit area. Thus 540 ft will be excluded and the tested range of Half_W_spcg will be from 660 ft to 1,320 ft
- Reducing the number of fracture stages in each well. Therefore, 210 ft will be excluded and the tested range of Frac_spcg will be from 420 ft to 1,050 ft
- Keeping Frac_cond constant at the middle level, 6 md.ft. This matches the results found in Cherian et al. (2013)
- Keeping Prod_BHP constant at the lower level, 1,600 psi. Lower pressure is not selected to avoid gas evolution around the wellbore if pressure goes below the bubble point pressure (1,569 psi), which would adversely affect oil production
- Keeping Frac_half_L constant at 300 ft to reduce the capital expenses

Table 4.2 summarizes the significant variables with the new ranges for the controllable ones. All the uncontrollable reservoir variables that were not significant are kept constant at their average level (Table 4.1).

Type	#	Variable	Unit	Levels			
				Low	Mid 1	Mid 2	High
Controllable Variables	1	Half well spacing (Half_W_spcg)	ft	660	880	1,056	1,320
	2	Frac spacing (Frac_spcg)	ft	420	630	840	1,050
Uncontrollable Reservoir Variables	3†	Initial reservoir pressure (P_res)	psia	5,000	6,000		7,000
		Reservoir temperature	deg F	225	240		255
		Depth	ft	9,500	10,500		11,500
	4	Rock compressibility (R_compr)	1E-6 psi^{-1}	3.50	6.00		8.50
	5	Matrix porosity (Poro)	%	2	6		10
	6	Matrix permeability (log(Perm))	md	0.0005 (-3.30103)	0.005 (-2.30103)		0.05 (-1.30103)
	7	Pay thickness (h)	ft	30	50		70
	8	Matrix endpoint relative permeability to oil at S_{wi} (K_{row})		0.1	0.55		1

†Only initial reservoir pressure was considered to avoid multicollinearity problems. Reservoir temperature and depth were modified accordingly.

Table 4.2—Optimization controllable and uncontrollable variables and their ranges.

4.3.2 Optimization

With the selection of eight significant variables, the next step is to fit a semi-empirical model to relate the response to these variables. A quadratic polynomial model is flexible and is the most widely used model in response surface methodology (RSM) applications because it can efficiently handle curved response functions which is often the case in optimization (Mathews 2005; Eriksson 2008; Montgomery 2012; Jamshidnezhad 2015). This quadratic model includes the linear, quadratic, and interaction terms as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \sum_{i=1}^{m} \beta_{ii} x_i^2 + \sum_{j<i} \sum_{i=2}^{m} \beta_{ij} x_i x_j + \varepsilon \qquad \text{(Eq. 4.6)}$$

where y is the objective function (response); x_i, i = 1, 2, …, m, are the variables; β_i (and β_{ij}), i = 1, 2, …, m, are the regression coefficients; m is the number of variables (eight in

the current case); and ε is the error. The summation term represents the quadratic terms and the double summation term represents the interaction terms. It should be recognized that machine learning techniques other than least squares regression could be used and probably give better results. However, the focus here is to illustrate the basic steps in least squares regression and how it can be verified using residuals diagnostics. More machine learning techniques will be used in Chapter 5 and their performance will be compared to that of least squares regression.

I-optimal design is chosen to generate this model because it outperforms the more commonly used D-optimal design since it minimizes the average variance of the estimates of model parameters and focuses on prediction. Thus it is more appropriate than the D-optimality criterion for RS designs (Jones and Goos 2012).

To permit accurate estimation of regression coefficients in the multi-linear regression model, the number of simulation runs needs to be determined. With eight variables, Eq. 4.7 could have as many as 45 coefficients. Thus the minimum number of simulation runs required must be at least 45. Austin and Steyerberg (2015) did a good literature regarding the number of runs per variable. Based on their work, the required number of runs is found to vary from 40 to 200. The number selected here is160 runs (20 runs per variable). G-efficiency of the generated I-optimal design is 0.635 which is within the recommended range provided by Eriksson (2008).

Fig. 4.9 shows boxplots of the NPV of the eight time steps for the 160 optimization runs. Most of the 160 runs have negative NPV even after 10 years of production.

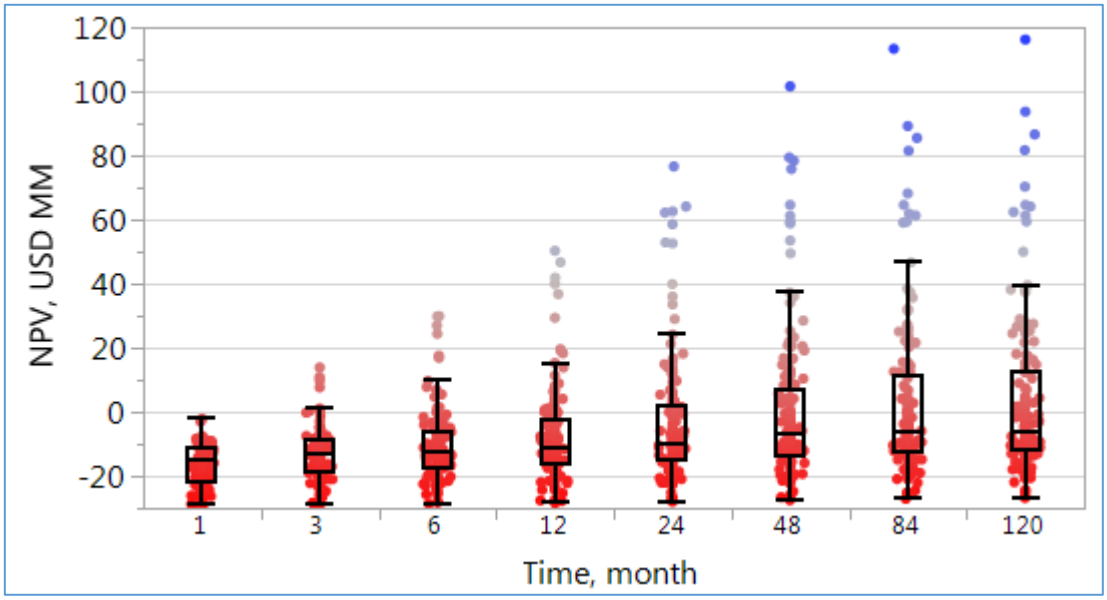


Fig. 4.9—Boxplots of the NPV of the eight time steps in the 160 optimization runs.

Fig. 4.10 shows how the two most significant variables (porosity and permeability, representing oil in place and flow rate, respectively) affect the 10-year NPV. The low levels of these two variables (2% and 0.0005 md) have negative mean NPV values. In addition, 55 out of the 63 runs with 2% porosity and 61 out of the 63 runs with 0.0005 md permeability have negative NPV. Thus it is evident that primary production from reservoirs with 0.0005 md permeability and/or 2% porosity is usually not economic. This suggests that these two variables might be better estimated in order to reduce the uncertainty in forecasting oil production. This also suggests that searching for sweet spots could be more economically beneficial than well and fracture design optimization. Investigating the potential of other recovery methods could be another solution to improve RF.

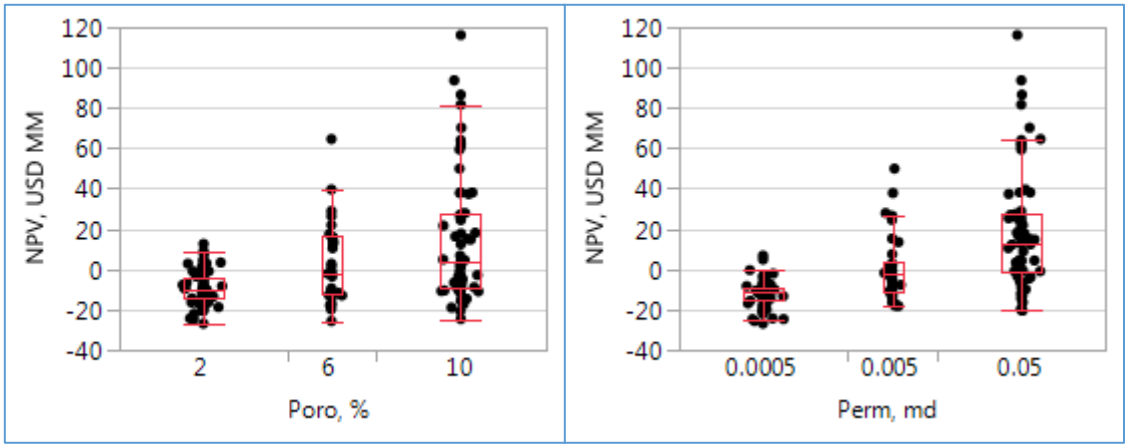


Fig. 4.10—Box-plots of 10-year NPV vs. the three levels of porosity and permeability.

4.3.2.1 Fitting RS

Mixed stepwise regression is performed to determine the best response surface (RS) based on model selection criteria (minimum root mean square error (RMSE), minimum predicted sum of squares (PRESS), minimum Mallow's Cp, minimum corrected Akaike Information Criterion (AICc), minimum Bayesian Information Criterion (BIC), and maximum adjusted coefficient of determination (R^2_{adj})). **Table 4.3** gives the estimates of regression coefficients of the 10-year NPV best RS. The objective NPV function (y in Eq. 4.6) is $(\boldsymbol{ln}(\sqrt{\boldsymbol{NPV + 50}}))^{\boldsymbol{-1.5}}$ with NPV in millions of dollars (USD MM).

The p-values of the regression coefficients are all less than the significance level, α = 0.05, indicating their significant contribution to the RS. Furthermore, power analysis is conducted using significance level, α = 0.05, and the value of power is given in the last column of Table 4.3. The resulting power of most model terms is more than 0.95, which means that in similar runs, there is a 95% chance of detecting a significant effect for the corresponding variables or interactions. This model has the following statistical properties:

RMSE = 0.008737, PRESS = 0.0155933, R^2 = 0.977731, R^2_{adj} = 0.972119, Cp = 31.6983 (< p = 33, number of predictors +1), AICc = -1012.71, and BIC = -927.196. Its analysis of variance (ANOVA) table is given in **Table 4.4**.

Term	Estimate	Prob>|t|	Power
Intercept	4.670E-01	<0.0001	
Half_W_spcg	-4.223E-05	<0.0001	1.000
h	-9.650E-04	<0.0001	1.000
log_Perm	-3.538E-02	<0.0001	1.000
P_res	-8.249E-06	<0.0001	1.000
Krow	-3.572E-02	<0.0001	1.000
Poro	-6.145E-03	<0.0001	1.000
R_compr	-1.644E-03	<0.0001	0.999
(Half_W_spcg-989.725)*(Half_W_spcg-989.725)	6.686E-08	<0.0001	0.986
(Half_W_spcg-989.725)*(Frac_spcg-737.625)	1.959E-08	0.0288	0.593
(Frac_spcg-737.625)*(Frac_spcg-737.625)	6.471E-08	0.0004	0.950
(Half_W_spcg-989.725)*(h-49.875)	7.543E-07	<0.0001	1.000
(Frac_spcg-737.625)*(h-49.875)	5.059E-07	0.0006	0.940
(h-49.875)*(h-49.875)	1.257E-05	0.0047	0.816
(Half_W_spcg-989.725)*(log_Perm+2.30103)	1.962E-05	<0.0001	1.000
(h-49.875)*(log_Perm+2.30103)	-2.600E-04	<0.0001	1.000
(log_Perm+2.30103)*(log_Perm+2.30103)	3.681E-03	0.0354	0.560
(Half_W_spcg-989.725)*(P_res-5975)	1.060E-08	0.0002	0.970
(Frac_spcg-737.625)*(P_res-5975)	1.037E-08	0.0004	0.949
(log_Perm+2.30103)*(P_res-5975)	-1.873E-06	0.036	0.557
(Half_W_spcg-989.725)*(Krow-0.54719)	2.464E-05	<0.0001	0.983
(h-49.875)*(Krow-0.54719)	-3.220E-04	0.0012	0.909
(log_Perm+2.30103)*(Krow-0.54719)	1.088E-02	<0.0001	1.000
(Krow-0.54719)*(Krow-0.54719)	3.461E-02	0.0001	0.978
(Half_W_spcg-989.725)*(Poro-6.025)	6.778E-06	<0.0001	1.000
(Frac_spcg-737.625)*(Poro-6.025)	5.431E-06	<0.0001	1.000
(h-49.875)*(Poro-6.025)	-3.466E-05	0.0019	0.883
(log_Perm+2.30103)*(Poro-6.025)	-3.360E-03	<0.0001	1.000
(Krow-0.54719)*(Poro-6.025)	-3.651E-03	<0.0001	1.000
(Poro-6.025)*(Poro-6.025)	4.824E-04	<0.0001	0.992
(Half_W_spcg-989.725)*(R_compr-5.95313)	2.700E-06	0.0148	0.689
(log_Perm+2.30103)*(R_compr-5.95313)	-1.104E-03	0.0019	0.881
(R_compr-5.95313)*(R_compr-5.95313)	6.668E-04	0.0182	0.661

Table 4.3—Estimates of RS regression coefficients, their p-values and power.

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Model	23	46.940526	2.04089	707.2793
Error	136	0.392435	0.00289	Prob > F
Total	159	47.332961		<.0001

Table 4.4—ANOVA table of the fitted Response Surface.

The actual by predicted plot is given in **Fig. 4.11**. It shows an acceptable match between the simulation output (actual) and RS predicted response (NPV function). The average absolute error between simulated NPV and RS predicted NPV is USD 2.766 MM.

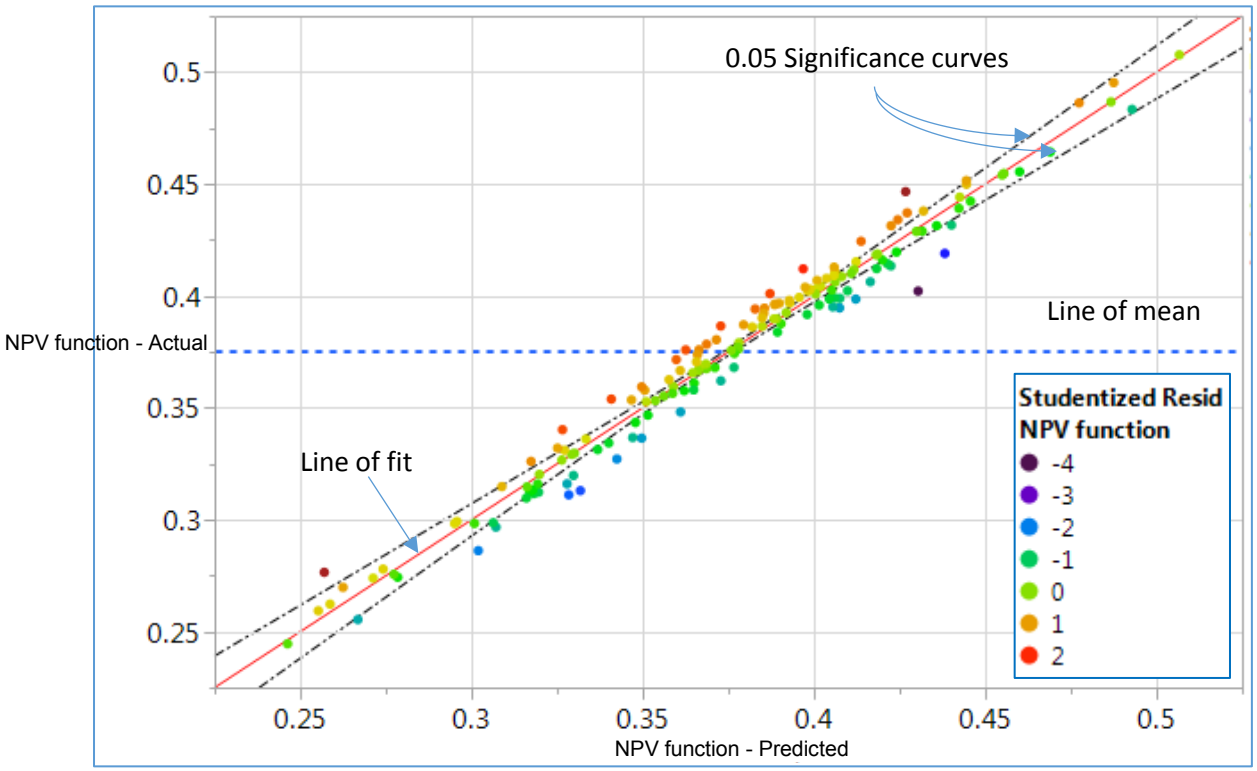


Fig. 4.11—Actual by predicted plot of the 10-year NPV response surface.

The model is checked further by Box-Cox plot (**Fig. 4.12**). It is a good diagnostic tool for identifying if data transformation is required. The current value of lambda (λ) in the model, one, lies within the recommended range that minimizes the residual sum of squares. Thus no transformation is required. This is an indication of model validity.

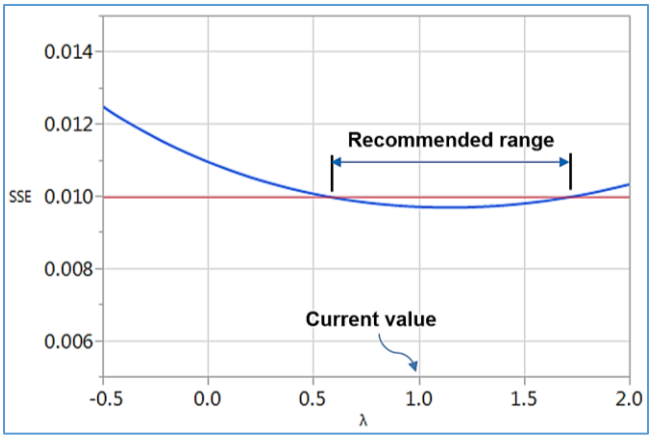


Fig. 4.12—Box-Cox plot.

Model diagnostics are then investigated to validate the model assumptions about the prediction errors (ε_i's) which are:

1. ε_i's are independent
2. ε_i's have constant mean 0 and constant variance σ^2
3. ε_i's have normal distribution

For independence, there should be no autocorrelation in the residuals (prediction errors). This is because the simulation runs would give the same results regardless of the order of doing them. Thus they should be independent. The second assumption is checked

using a plot of the studentized (standardized) residuals versus the predicted model responses (**Fig. 4.13**).

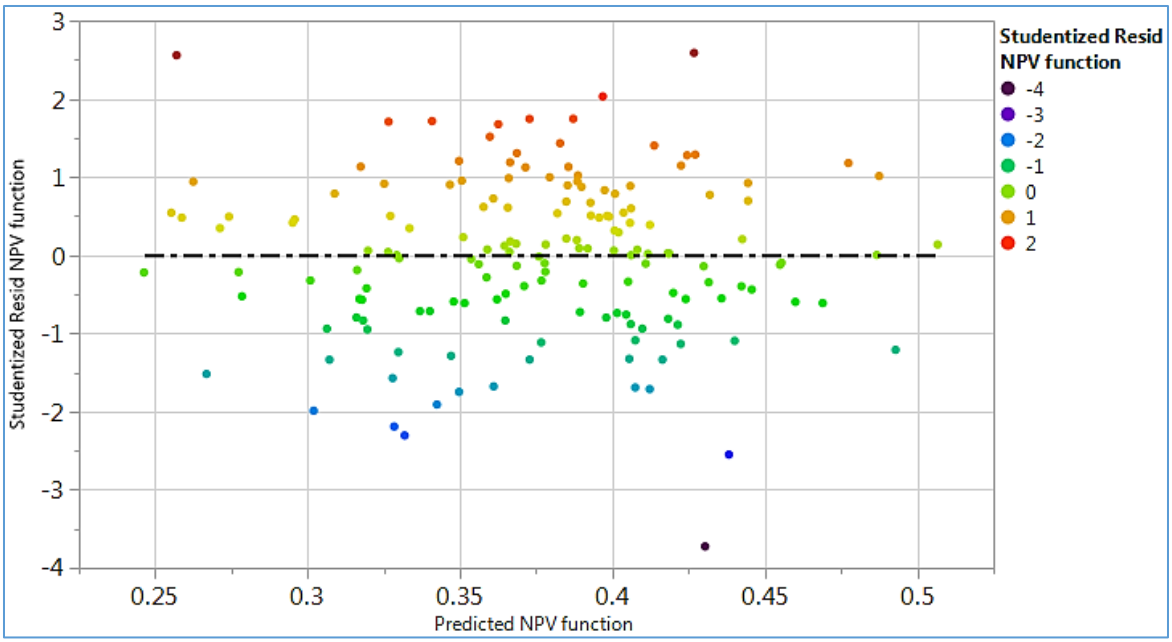


Fig. 4.13—Studentized (standardized) residual by predicted plot for NPV function.

As it can be seen, the scatterplot looks flat validating the constant zero mean assumption. In addition, the vertical scatter of the data does not change horizontally, indicating the validity of constant variance assumption. The same conclusions can be derived from the analysis of the residual plots of the studentized residuals versus each of the eight predictors (**Fig. 4.14**).

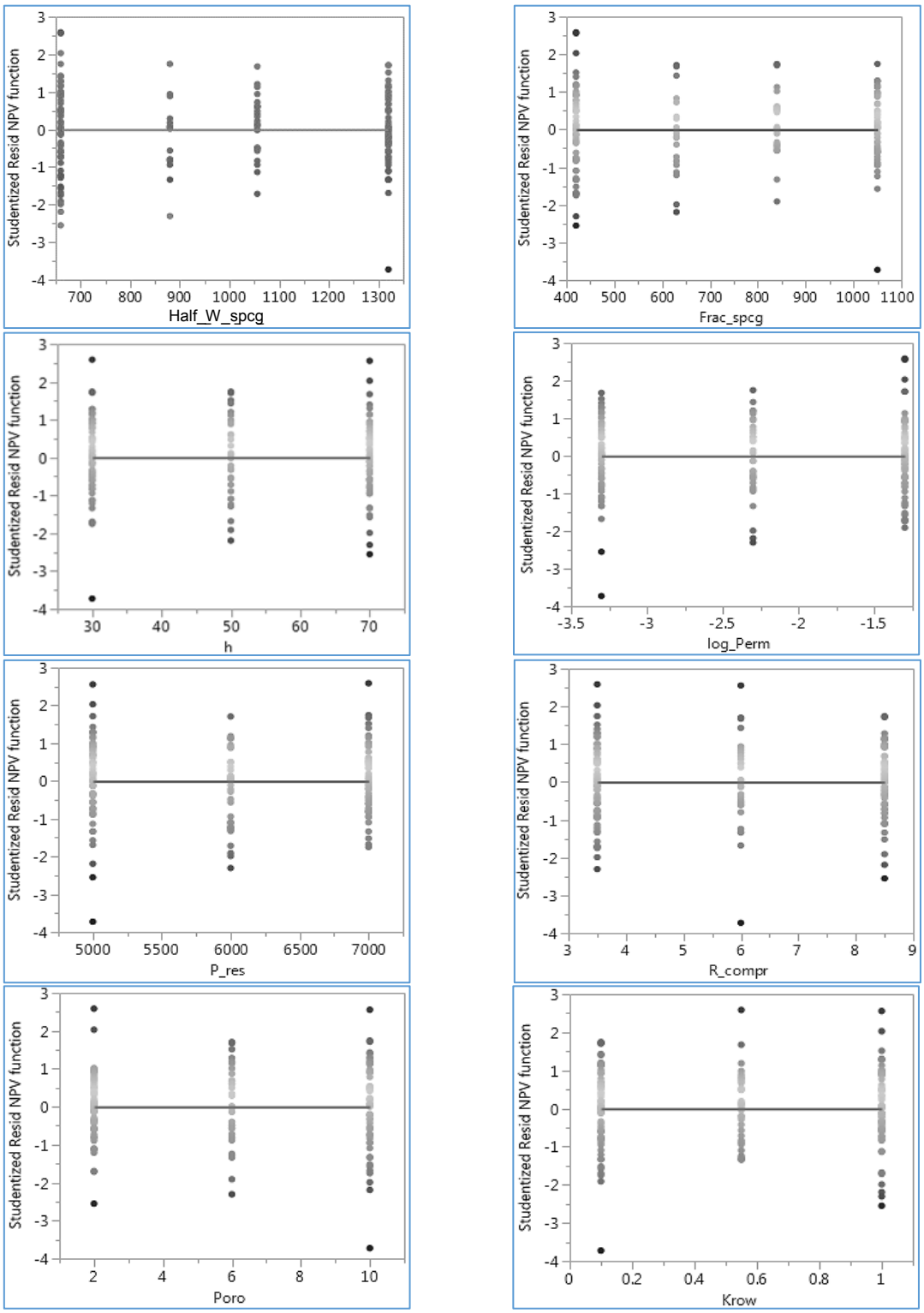


Fig. 4.14—Plots of studentized residual by the eight significant variables.

The third and last assumption is checked by taking a closer look at the observations that could affect the normality assumption. These observations could be unusually large residuals (outliers) or data that make exceptional contribution to the model fit or analysis (influential observations). Outliers are detected by calculating studentized residuals for each observation. Its absolute value should be 2.5 or greater to indicate a potential outlier. In the present case, three observations have absolute values greater than 2.5 (2.58839, 2.55862, and -3.72404). However, this can be considered acceptable for 160 observations. For influential observations, Cook's D Influence should be 0.2 or larger to consider the observation as influential. In this case, none of the Cook's D Influences is beyond 0.2 and, therefore, no influential observations exist. These conclusions validate the normality assumption.

The normality assumption is also checked by considering the studentized residuals' normal quantile plot and histogram (**Fig. 4.15**). The distribution histogram of the residuals seems normal with one outlier and a normal curve fits it well. The normal quantile plot shows that all the points lie between the 95% confidence limits along the diagonal reference line validating the normality assumption. Additionally, a normality hypothesis test (Shapiro-Wilk W test) gives the same conclusion as shown below.

H_0: data is from normal distribution; H_1: data is not from normal distribution (Eq. 4.8)

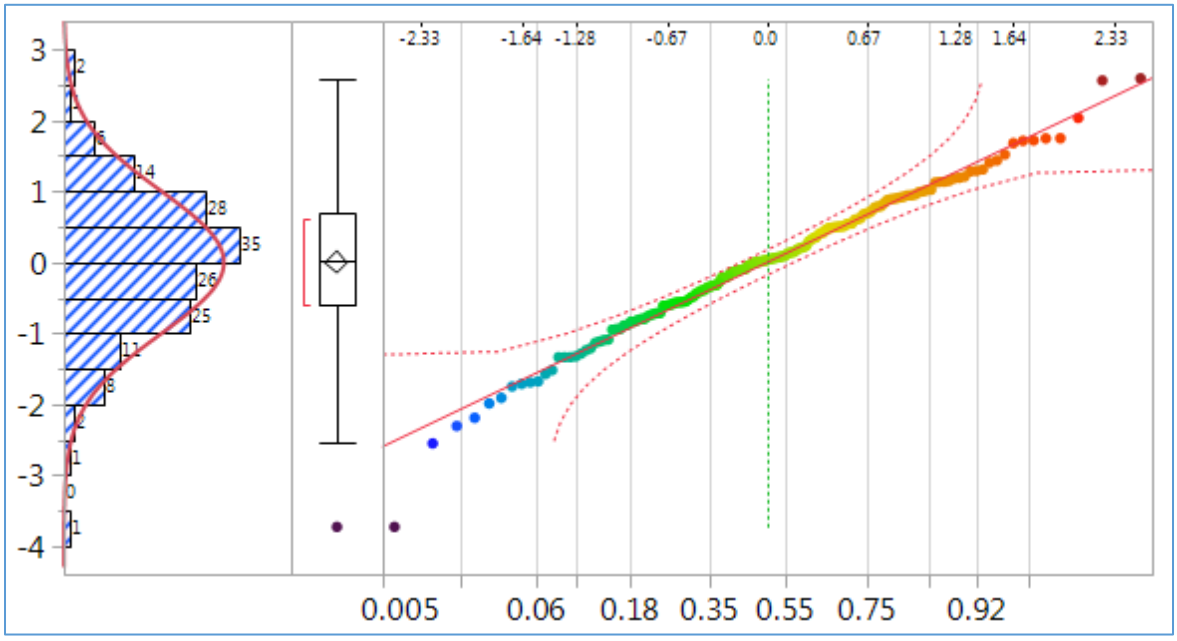


Fig. 4.15—Normal quantile plot and distribution of studentized residuals of NPV function with normal fit.

The Shapiro-Wilk W statistic is 0.990541 and the p-value is 0.3654. With significance level, $\alpha = 0.05$, the null hypothesis (H_0) cannot be rejected because p-value > α, and there is no strong evidence against the normal distribution hypothesis. Now the model is ready for further use.

4.3.2.2 Optimum Operating Conditions

The NPV function model is then used to find the optimum combination of significant controllable variables (Half_W_spcg and Frac_spcg) at the average level of the uncontrollable reservoir variables, i.e., log_Perm = -2.30103 (Perm = 0.005 md), P_res = 6,000 psi, h = 50 ft, Poro = 6%, Krow = 0.55, and R_compr = 6×10^{-6}psi^{-1}. Mathematical optimization yields optimum Half_W_spcg = 1,310 ft and Frac_spcg = 688 ft. The NPV for this case is USD 12.8 MM using the RS and USD 6.4 MM using simulation. The

prediction accuracy could be increased by using the techniques in Chapter 5. However, the RS is still considered reliable and is used in further analysis. **Fig. 4.16** shows the cumulative oil production and cumulative NPV over the production period. The payback time is around 3.2 years.

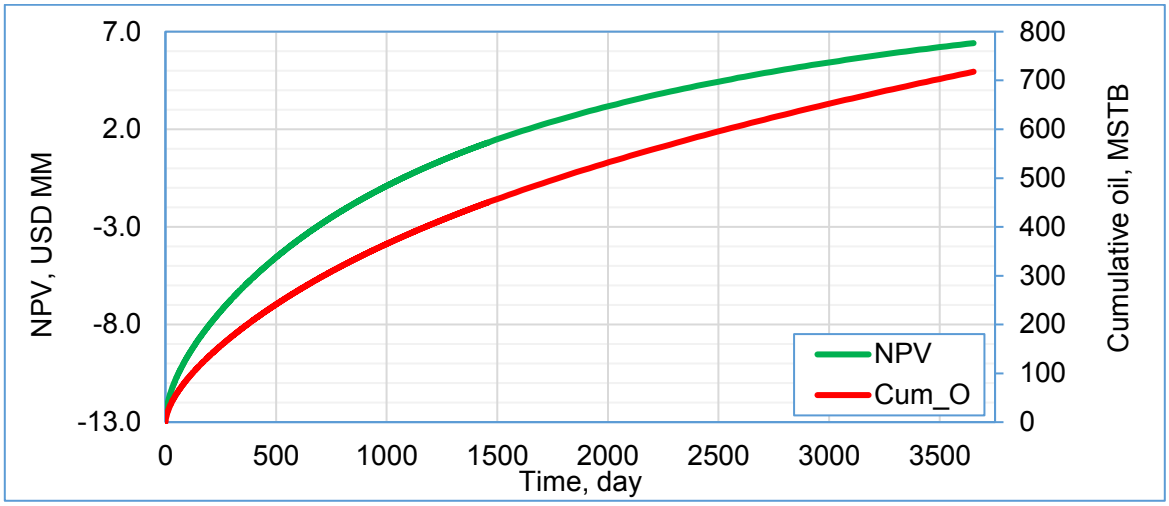


Fig. 4.16—Economic performance of the optimum run.

The effects of the less significant design variables are then investigated. Producing bottom-hole pressure is not investigated because it is kept constant at 1,600 psi, the lowest possible pressure to avoid the evolution of gas. For fracture conductivity and half-length, two more levels are investigated: 120 ft and 480 ft for fracture half-length and 0.6 and 60 md.ft for fracture conductivity. **Table 4.5** summarizes the results of the simulation runs. Increasing both fracture conductivity and fracture half-length from their base-case values leads to increasing the NPV by around USD 5.2 MM. This clearly reveals that considering the variables that were screened out by the screening study is vital to maximize the NPV.

Case	Base	120 ft/ 0.6 md.ft	120 ft/ 60 md.ft	480 ft/ 0.6 md.ft	480 ft/60 md.ft
NPV, USD MM	6.41	-2.3	-0.6	2.5	11.6

Table 4.5—Effects of the less significant design variables.

4.3.2.3 Uncertainty Analysis (Monte Carlo Simulation)

The final step is to conduct Monte Carlo (MC) simulation to quantify the uncertainty in the generated NPV. Using the response surface, MC uncertainty analysis is much simpler than stochastic methods and is almost as accurate as MC analysis with fine-scale simulation. The distribution of the various variables is assumed as follows: uniform for both Half_W_spcg and Frac_spcg; log-normal for permeability (i.e., normal for log_Perm); and triangular for P_res, h, Krow, log_Perm, R_compr, and Poro. The average value (Table 4.2) of the uncontrollable state variables is used for triangular distributions. The number of iterations (n) conducted is 50,000. The distribution of the NPV is given in **Fig. 4.17**. The median value is USD 4.93 MM. 62.53% of the runs produced positive NPV. 78.73% of these runs have permeabilities larger than the average Bakken permeability, 0.005 md (Sarg 2012). 98.18% of the remaining 37.47% runs (runs with negative NPV) have permeabilities lower than 0.005 md. This suggests that primary production is mostly uneconomic for Bakken reservoirs with permeabilities lower than 0.005 md. Exploring other recovery methods is a necessity to boost the economics of Bakken shale oil production with permeabilities lower than 0.005 md. Also searching for sweet spots rather than focusing on well and fracture design optimization could be more rewarding.

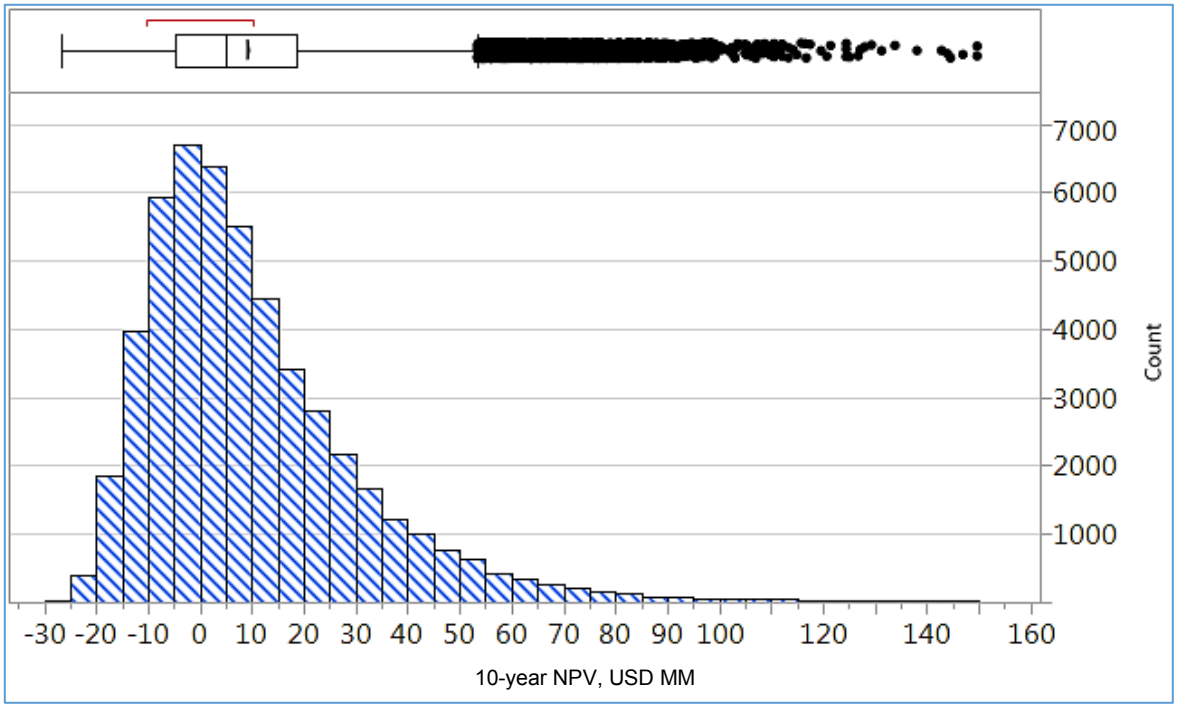


Fig. 4.17—10-year NPV probability distribution for 50,000 Monte-Carlo iterations.

4.4 Conclusions

Using simulation and design of experiments (DoE), this chapter deals with primary production (PP) in a heterogeneous Bakken shale reservoir. RSM was applied and verified using residuals analysis. The objectives are to determine the most significant variables for the 10-year NPV, compare the relative importance of controllable design variables and uncontrollable reservoir characteristics, optimize PP design for maximum NPV, and quantify the uncertainty in NPV. The main findings and conclusions of the study are as follows:

1. Six out of the eight statistically significant variables are reservoir characteristics indicating their dominant influence on reservoir economic performance. PP is

mostly uneconomic for Bakken reservoirs with permeabilities lower than 0.005 md. Priority should be given to searching for sweet reservoir spots and EOR rather than drilling and completion optimization.

2. Being the two most significant variables, porosity and permeability should be better estimated in order to reduce the uncertainty in forecasting oil production.
3. Design variables initially dominate the effect on production economics, and reservoir characteristics take over later (after one to two years) in the life of the well.
4. For average Bakken shale properties, the optimum conditions are 1,310 ft and 688 ft for half well spacing and fracture spacing, respectively. The 10-year NPV for this case is USD 6.4 MM. However, if the screened-out design variables (fracture conductivity and fracture half-length) are considered, the NPV jumps to USD 11.6 MM.

5. REDUCING BIAS BY DATA ANALYTICS AND STATISTICAL ANALYSIS: APPLICATION ON BAKKEN SHALE

5.1 Introduction

Statistical design of experiments (DoE) is generally overlooked in the petroleum industry (Chapter 2). The same is true for data analytics (DA) (Schuetter and Mishra 2015). **Fig. 5.1** shows the trend of the appearance of some keywords related to DA in SPE online library (onepetro.org) until the end of 2016. Despite the continuous increase with time, more efforts should be directed towards increasing the awareness of such techniques. This is especially true for DoE, which should become the standard methodology of conducting analytical research.

The lack of use of such techniques could deteriorate research efficiency and lead to biased conclusions stemmed from reduced accuracy and precision of results. Such a problem could be especially aggravated in complex and marginally economic reservoirs like shale, where large uncertainties intensify the risk in the decision-making process. Therefore, Chapters 3 and 4 employed DoE and statistical analysis using reservoir simulation to understand shale oil recovery process and improve its low recovery factor (RF).

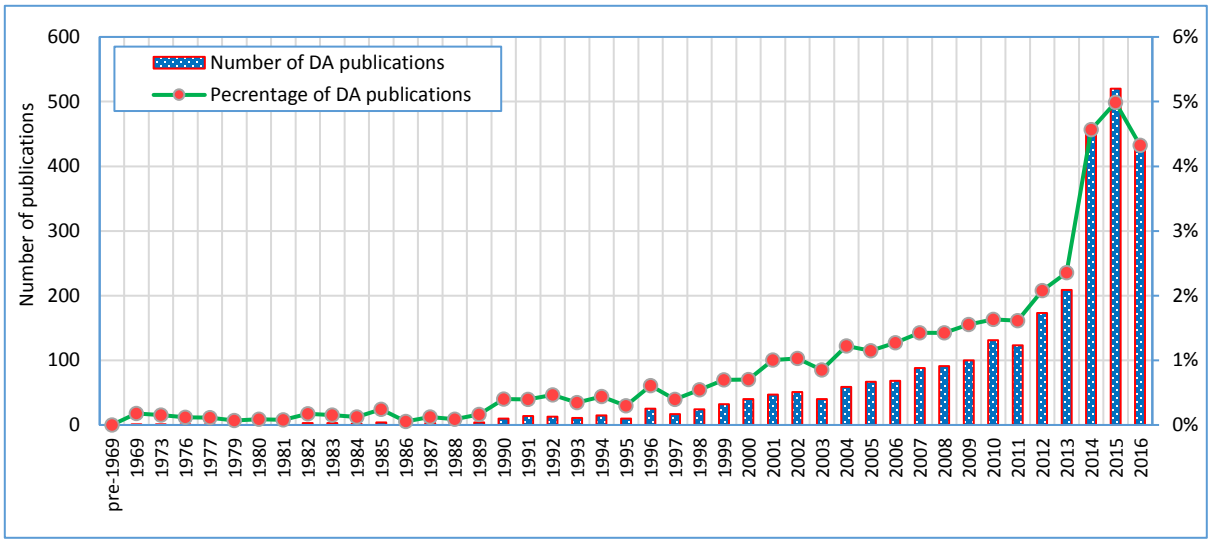


Fig. 5.1—Numbers and percentage of DA papers.[2]

However, adequate representation of the complex and challenging shale reservoirs using numerical simulation necessitates making numerous uncertain assumptions. In addition, these techniques could be time consuming and their efficiency deteriorates in case of complicated and highly non-linear response-predictors relationships. Thus other less presumptive, more effective, and faster methods are required to support modeling results (Mohaghegh et al. 2017). DA offers a powerful tool in this respect because it avoids making the physical assumptions associated with simulation. Thus it reduces the

[2] DA search keywords are "data mining", "machine learning", "big data", "data analytics", "predictive analytics", "predictive modeling", "data science", "supervised learning", and "unsupervised learning" (search done July 8th 2017). Note that only years with DA publications are shown and that mentioning these keywords does not imply that DA was actually implemented.

uncertainties associated with shale oil production, which could reduce risks and improve the robustness to the decision-making process. Furthermore, it could be faster if data preparation is not a big issue, and more effective for complicated response-predictors relationships.

This chapter harnesses the evolutionary potentials of data analytics and applies it to a big dataset for Bakken shale to produce statistically-based conclusions and data-driven facts. Several authors applied data analytics on Bakken shale (Lafollette et al. 2012; Izadi et al. 2013; Ling et al. 2016; Lolon et al. 2016; Wang and Chen 2016b). However, not many predictive analytics techniques were employed. In addition, the results sometimes disagree for different techniques and studies. Furthermore, some data sets are relatively small. Thus this study uses a more comprehensive data analytics (descriptive and predictive) framework on an 8,095-well dataset by applying and comparing the results of 12 supervised learning (SL) techniques for four production performance metrics. These techniques are artificial neural networks (ANN), regression trees (RT), bootstrap (or random) forests (BF), boosted trees (BT), k-nearest neighbors (KNN), linear regression (LR), ridge regression (RR), least absolute shrinkage and selection operator (LASSO), adaptive LASSO (A-LASSO), elastic net (EN), adaptive EN (A-EN), and finally ensembles (ENS). The objectives of this work are to:

1. Compare the predictive performance of the 12 SL techniques for four production performance metrics
2. Identify key factors that distinguish good wells from poor-performing ones

3. Mine data-driven insights from the available big data to better understand the complex relationships and interactions between drilling, completion, and reservoir data on one side, and production data on the other side
4. Predict the production performance of future wells based on their location, completion design, and well architecture
5. Evaluate the validity of the results of Chapter 4 to assess the performance of simulation

5.2 Methodology

JMP statistical software is used here to do statistical analysis.

5.2.1 Data Preparation, Exploration, and Reduction

This step incorporates domain knowledge, data summaries, data conversion, dimension reduction, regression imputation, multivariate normal imputation, and univariate and bivariate analytic techniques to understand the overall structure of the dataset and reduce its dimensionality when required. The dataset includes drilling, completion, and production data of 19,109 wells in Bakken shale play. There are 5,449 wells with no production data, 95% of which also have no completion data at all. Thus these wells were deleted. Less than 3% of the 3,556 wells completed before 2011 have completion data. These wells were discarded as well. Vertical wells were also removed from the dataset to focus on horizontal wells. If all such wells were not removed, more than 50% of the completion predictors would have been missing, which could reduce the quality of study findings and conclusions. Boxplots, scatterplots, cross-plots, and color-coded maps were used to clean the data and explore the response-predictors relationships.

Further cleaning of the remaining dataset continued to replace or delete values that did not make sense, and to impute missing values. For example, regression analysis was used to fit a linear relationship between completion date and first production date with determination coefficient (R^2) of 0.99935. This relationship was then used to impute missing completion dates. Completion date was used to create a new variable, Quarter (wells completed in the first quarter of 2011 have Quarter = 1). The analysis left nine predictors to be used in this work. They included eight continuous predictors as follows: well location (Latitude and Longitude), Quarter, Vertical Depth (foot or ft), Lateral Length (ft), amounts of Water and Proppant used for fracturing (US gallon (gal) and pounds (lb), respectively), and number of Fracture Stages. The ninth predictor is Pad ID and it is categorical. Wells are considered to have Pad ID = 1 if they are located less than 250 ft apart.

Four production metrics are used as responses. They are IP, EUR, Cum90, and Cum365, which represent initial production (b/d), estimated ultimate recovery (million bbl or mmbbl), cumulative 90-day production (bbl), and cumulative 365-day production (bbl), respectively. EUR is estimated based on 30-year decline curve analysis. The production of the first month (Cum30) usually does not represent production over a 30-day period because production might not start at the beginning of the month. In addition, early production could be subject to early changes. Thus Cum30 was not considered as one of the production performance metrics. The four metrics were predicted using 12 SL techniques to strengthen the reliability of the results because incorrect values most likely still remain in the dataset despite the tedious data cleaning and preparation. Another reason

for using four different responses representing different production times is to assess the performance of simulation of Chapters 3 and 4 and to provide better representation of well performance on the short as well as long terms.

Further imputation of missing values of other predictors proceeded. Multivariate normal imputation - a multiple imputation (MI) method - was used to impute the missing values based on a multivariate normal distribution. A shrinkage estimator is used to enhance the covariance matrix estimation (Klimberg and McCullough 2016). MI assumes that the missing data mechanism is ignorable (missing completely at random, MCAR or missing at random, MAR). Although the current dataset may not satisfy this assumption and the missing data mechanism could be missing not at random or MNAR (**Fig. 5.2**), Schafer and Graham (2002) found that MI methods are often unbiased with NMAR data. Multivariate normal imputation was also applied to impute the missing values for the four production metrics. However, imputation considered only various other production metrics, which are highly correlated to the four chosen metrics. No predictors were used here for this imputation.

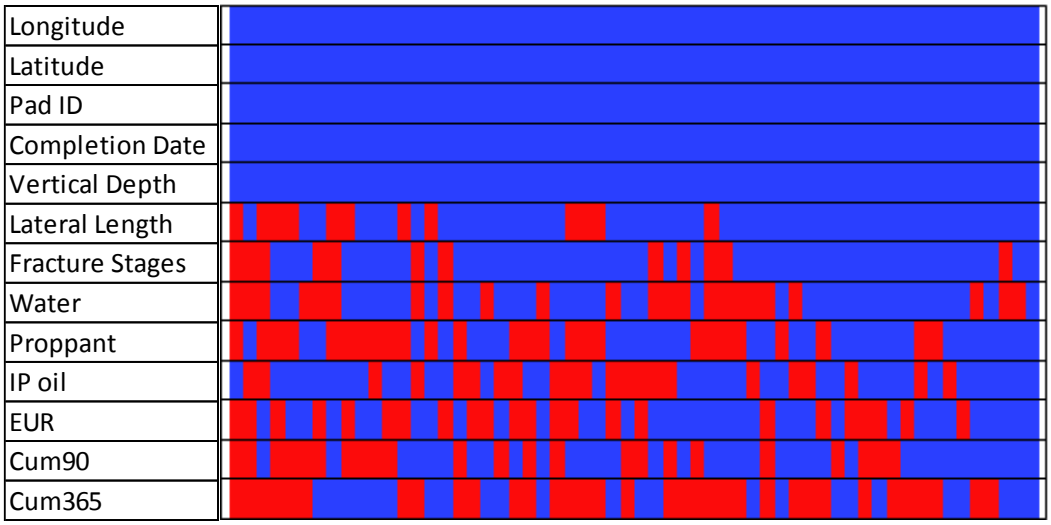


Fig. 5.2—Missingness pattern cell plot of the nine predictors and four responses.

Extreme outliers of the predictors were removed using Mahalanobis distance. Euclidean distance was not used because it depends on the units used for predictors, does not consider variability of predictors, and ignores the correlations between them. Mahalanobis distance offers an effective solution to such problems (Shmueli et al. 2017). **Fig. 5.3** shows Mahalanobis distance plot with alpha of 0.001. Wells that have Mahalanobis distance above an upper control limit (UCL) of 5.3 are considered outliers and removed from the dataset. The final complete dataset had 8,095 wells.

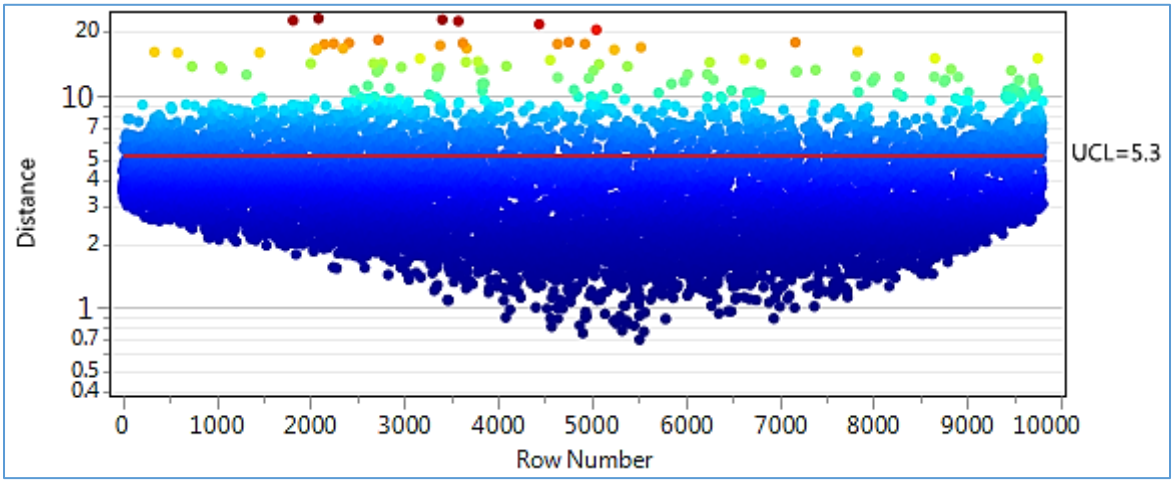


Fig. 5.3—Mahalanobis distance plot with alpha of 0.001.

5.2.2 Model Evaluation and Selection

Several metrics were used to assess the fit and compare the predictive ability of various models. These measures include root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and median absolute error (MdAE). RMSE represents the standard deviation of the differences between the actual and predicted values. R^2 represents the proportion of variability in the response that is predicted by the predictors. MAE gives an idea about the central tendency of the errors. Absolute error is used because negative errors would cancel out positive ones, resulting in a mean value that is not truly indicative of the errors. MdAE is used because MAE could be highly influenced by outliers in the errors. In addition, since absolute errors disregard the magnitude of the variable, some relative measures of error are also considered. These include mean absolute relative errors (MARE), median absolute relative errors (MdARE), and relative squared error (RSE). These metrics are defined by the following equations:

Coefficient of determination (R^2) = $1 - \sum_{i=1}^{n}(\hat{y}_i - y_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2$ (Eq. 5.1)

Root mean square error (RMSE) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$ (Eq. 5.2)

Mean absolute error (MAE) = $\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$ (Eq. 5.3)

Mean absolute relative errors (MARE) = $\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|/y_i * 100$ (Eq. 5.4)

Relative squared error (RSE) = $\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2$ (Eq. 5.5)

where $\hat{y}_i$, y_i, and $\bar{y}$ are the i^{th} estimated response, i^{th} actual response, and average of n actual responses, respectively. n is the number of data points in the training or validation sets. The two median measures are self-explanatory.

To avoid overfitting, the 8,095-well dataset was split into three sets. The first is a training set used to train the 12 SL methods. The second set is for validation to guard against overfitting by comparing the evaluation metrics of the validation set to their corresponding values in the training set. An increase in the difference between the metrics of both sets indicates overfitting and using complex models. This happens because the error in the validation set starts to increase after a certain level of training at which the model starts to fit the random errors instead of the underlying relationship. In addition, the evaluation metrics of the validation set are used to compare the predictive performance of all generated models in each of the 12 SL techniques to choose the best model for each technique. Then, the validation set is also used to compare the predictive performance of the 12 best models to determine the superior model that outperforms all others. However, this "superior" model could still be overly optimistic. Thus a test set is used after training and validation to provide an unbiased measure of the performance of the best model. The

test set accounts for 25% of the dataset. The remaining set is divided into 75% for training and 25% for validation.

5.2.3 Supervised Learning (SL)

Here various SL methods are applied to produce models for various production performance metrics. Twelve methods were used and their predictive performance for the four responses were compared together. In addition, the contribution of the nine predictors to the four production responses were considered. These 12 methods are described below (Klimberg and McCullough 2016; Shmueli et al. 2017).

1. Artificial neural networks (ANN): Only one hidden layer is used. The hyperbolic tangent transformation is used as the activation function for the hidden nodes and a linear combination of the generated features from the hidden nodes is used for the four output nodes corresponding to the four responses. Several numbers of hidden nodes were tried to search for the numbers that best model the four responses without overfitting. The strategy started with a minimum number of three nodes and continued to increase progressively until the prediction error of the validation (hold-out) set started to increase. Up to 40 nodes were tried. Care was considered to avoid overfitting, which was indicated by a training error that is much higher than the validation error. Several techniques were employed to improve the fit. First, the continuous variables were transformed to near normality because they are skewed and have outliers. Second, gradient boosting was employed to enhance the predictive performance of the ANN. The number of models tried in boosting was 20, and a low learning rate of 0.1 was used to reduce the chance of overfitting.

Another technique used to avoid overfitting was the penalty method, which penalizes the square of the estimated weights. The fourth and final technique used to reduce the chance of getting trapped in a local optimum was to restart the ANN model several times using various sets random initial weights, and selecting the best model. Twenty restarts were used as recommended by Sall et al. (2007).

2. Regression trees (RT): The number of splits was varied for each response while keeping an eye on R^2 and RMSE for both training and validation sets to avoid overfitting.
3. Bootstrap forests (BF): To get the best BF model, the number of trees in the forest was varied up to 1,000, the number of variables sampled at each split was varied from two to nine, and the number of splits per tree ranged from 10 to 2,000. The minimum number of instances used for a node to split is five as recommended by Hastie et al. (2009). Adding more trees was stopped when it did not improve the validation R^2.
4. Boosted trees (BT): To get the best BT model, the number of trees that could be considered was up to 1,000, the number of splits varied from three to 100, and learning rate varied from 0.1 to 0.9. As shall be seen later, the four best models for the four responses have a learning rate of 0.1. The minimum number of instances used for a node to split is five.
5. K-nearest neighbors (KNN): The predictors are standardized to eliminate the effect of different units of measurement. The value of k was varied from 1 to 50 and the

value that provides the model with-the least validation RMSE was chosen for each response.

6. Multiple linear regression (LR): Mixed stepwise regression was employed with a p-value of 0.05 for the variables to enter or leave the model. A response surface model that has linear, quadratic, and interaction terms was considered. Overall F-test for the whole model as well as individual t-tests for each predictor were conducted. FDR (false discovery rate) LogWorth was considered to assess variables' importance. FDR LogWorth modifies the p-value in such a way that considers the false discovery rate associated with multiple tests. Any variable whose FDR p-value exceeded 0.05 was discarded. Tens of models were investigated for each response to make sure that regression assumptions are met. The assumptions and how they were treated are as follows:
    - Linear relationship: Residual plots were ensured to have constant zero mean.
    - Homoscedasticity (constant variance): At first, all the residual plots were fan shaped indicating non-constant variance. Thus Box-Cox transformations were used to transform the response variables to help eliminate such behavior of the residuals.
    - Normality: Normal quantile plots for standardized residuals were used to check the normality of residuals. No problems were detected.
    - No auto-correlation: Durbin Watson test was applied to check for autocorrelation. No problems were found as the residuals were not serially auto-correlated.

In addition, multicollinearity was considered. Variables with variance inflation factor (VIF) above 6.0 were discarded. Eliminating such variables improved the stability of the estimates of the regression coefficients. This agrees with the principle of parsimony which states that better models have less variables.

To reduce prediction error of LR, a penalty was applied to large coefficients. This was done by using a shrinkage penalty, which introduces some bias to the criterion of minimizing the residual sum of squares (RSS) of LR. This bias could reduce the variance, which translates into an improvement in prediction ability of LR (bias-variance trade-off).

Three penalization methods were considered here: ridge regression (RR), least absolute shrinkage and selection operator (LASSO), and elastic net (EN). In addition, the adaptive versions of the last two methods (adaptive LASSO (A-LASSO) and adaptive EN (A-EN)) were used as well. These adaptive versions improve the fit by reducing the penalty applied to variables that are more influential to the prediction. These five versions of generalized linear regression (GLR) are considered as separate methods (seven to 11).

7. Ensemble (ENS): This is the 12^{th} method. In an ensemble, the 11 SL models are combined together into a supermodel. This could reduce the variance of the errors and therefore, improve the predictive performance by providing more precise predictions. Median, average, weighted average and best-fit ensembles were investigated. The last two were based on the following form:

$$\hat{y}_{ENS} = \sum_{i=1}^{11} c_i * \hat{y}_i \quad \text{(Eq. 5.6)}$$

where $\hat{y}_{ENS}$ is the estimated ensemble value for any of the four responses for any well; i is a count that varies from 1 to 11 and accounts for the 11 SL techniques; c_i is the coefficient (best fit) or weight (weighted average) of the i^{th} technique; and $\hat{y}_i$ is the predicted value of any of the responses with the i^{th} technique.

For weighted average and best-fit, a non-linear optimization problem was set up for each of the four responses. The weights/coefficients of the 11 SL techniques were varied with the purpose of minimizing the objective function, RMSE of the training set. Central finite difference was used in optimization because it gave more accurate results. The process was repeated using different random starting sets of values of the weights (250 sets) to enhance the chance of finding the global optimum. Then error analysis was conducted. RMSE and R^2 of both training and validation sets were compared to avoid overfitting.

5.2.4 Importance of Variables

The contributions of the predictors to the four responses were evaluated. For LR, the relative importance of each predictor is measured by its standardized beta coefficient. Solution path plots were used for the five GLR techniques except RR. RR did not produce discernible solution path plots and therefore, the contributions of the predictors were not clear. The same is true for EN for all responses except EUR. For ANN, the importance of variables was evaluated based on the variability of the predicted response that results from changing the predictors in their ranges. If variation of a predictor resulted in high variability in the predicted response, the effect of such predictor is considered important.

For RT, BT, and BF, the ranking of a predictor's importance was based on its contribution to the explanation of the variance of the response. KNN and ENS do not produce such contributions (Klimberg and McCullough 2016; SAS Institute Inc. 2016; Shmueli et al. 2017). Interactions were also considered in the six regression techniques.

5.3 Results and Discussion

5.3.1 Data Exploration

Fig. 5.4 shows the distribution of the wells color-coded based on Cum365. The distribution is not random. There are some sweet spots which have higher productivity on average. The counties with the highest well densities are McKenzie, Mountrail, Williams, and Dunn.

Fig. 5.5 displays a scatterplot for the nine predictors along with their distributions. No strong correlations exist between any couple of predictors except for Vertical Depth and Latitude (0.6288). This is clearer in **Table 5.1**. This was considered in LR because of the multicollinearity it introduces into the models. **Fig. 5.6** focuses on the four highest productive counties and shows how completion predictors have evolved through time and their effect on Cum365 (similar effects on the other three production metrics were observed). Lateral Length is almost constant (9,000 to 10,000 ft) with time. Stages increase with time. Because Lateral Length is almost constant, Stages/1,000 ft also increase with time. Water, Water/stage, Proppant, and Proppant/stage increase with time. On the other hand, Proppant Concentration decreases slightly with time on average. Fig. 5.6 also shows that Cum365 increases with time. This is particularly clearer for McKenzie and Dunn. This could be due to the continuous increase of the amount of Proppant and Water per stage and

decrease of fracture spacing. It must be stated that the purpose of showing trends in cross-plots, like Fig. 5.6, is to demonstrate whether there is a positive or negative trend between the presented variables. That does not imply that there is a linear relationship (correlation does not imply causation).

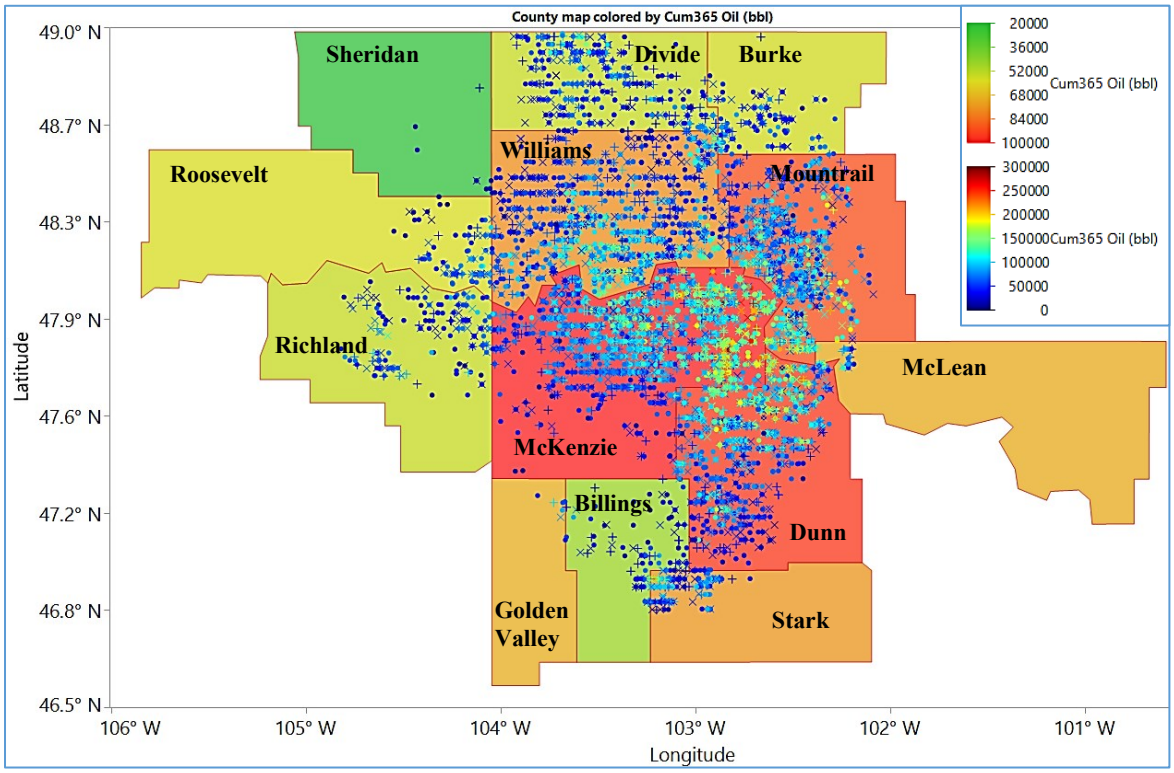


Fig. 5.4—County map for Bakken play wells color-coded based on Cum365. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

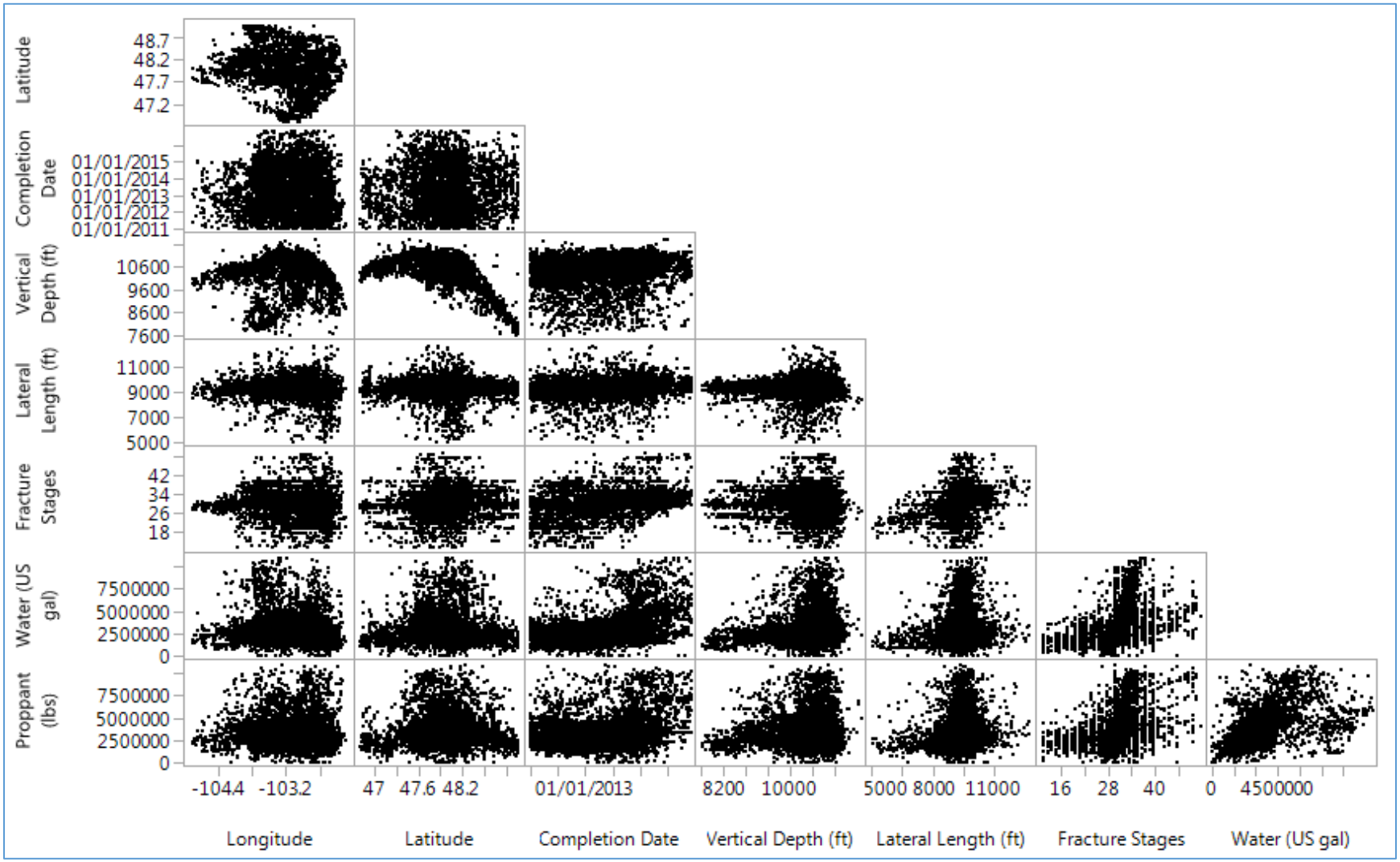


Fig. 5.5—Scatterplot matrix for nine predictors. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

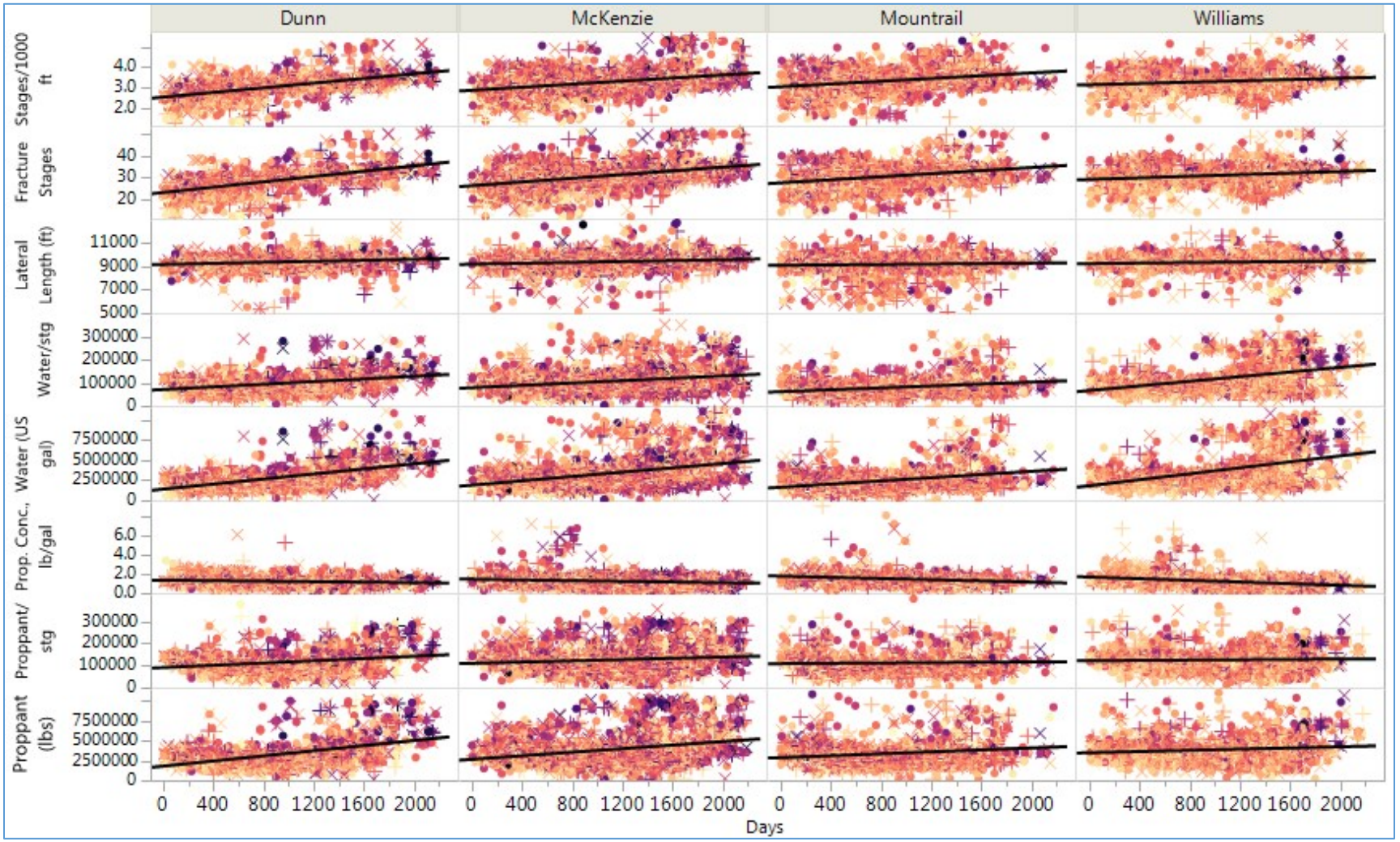


Fig. 5.6—Trellis plot of completion predictors vs. Days (time since January 1st 2011). Darker points mean higher Cum365. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

Predictor	Longit-ude	Latit-ude	Compl Date	Vert. Depth (ft)	Lateral Length (ft)	Frac. Stages	Water (US gal)	Proppant (lbs)
Longitude	1.0000	-0.1679	0.0302	-0.0454	-0.0417	-0.0249	-0.1235	0.0150
Latitude	-0.1679	1.0000	0.0177	-0.6288	-0.0388	0.0885	0.0216	0.0261
Completion Date	0.0302	0.0177	1.0000	0.1073	0.1221	0.3616	0.4273	0.2605
Vertical Depth (ft)	-0.0454	-0.6288	0.1073	1.0000	0.0658	0.0195	0.2009	0.1220
Lateral Length (ft)	-0.0417	-0.0388	0.1221	0.0658	1.0000	0.2797	0.1575	0.1360
Stages	-0.0249	0.0885	0.3616	0.0195	0.2797	1.0000	0.3509	0.3958
Water (US gal)	-0.1235	0.0216	0.4273	0.2009	0.1575	0.3509	1.0000	0.5558
Proppant (lbs)	0.0150	0.0261	0.2605	0.1220	0.1360	0.3958	0.5558	1.0000

Table 5.1—Correlation matrix of the nine predictors.

Fig. 5.7 shows that production metrics normalized with respect to Lateral Length increase with reducing fracture spacing (increasing number of stages per 1,000 ft lateral). Recently, the average number of stages per 1,000 of lateral has been between 3.5 and 5.5. The Lateral Length-normalized production metrics also increase with increasing Proppant and Water per stage and slightly decrease (except Cum90) with Proppant Concentration. The top 10% performing wells based on Cum365 tend to have more Water and Proppant per stage, 300 ft for fracture spacing, and 1 – 2 lb/gal Proppant Concentration. Another important observation is that the normalized EUR fitted line vs. Stages/1,000 lateral is the flattest among the four normalized responses. This implies that the effect of reducing fracture spacing is least on normalized EUR. Reducing fracture spacing has more to do with accelerating oil flow than improving the ultimate recovery, which agrees with the findings of Chapters 3 and 4.

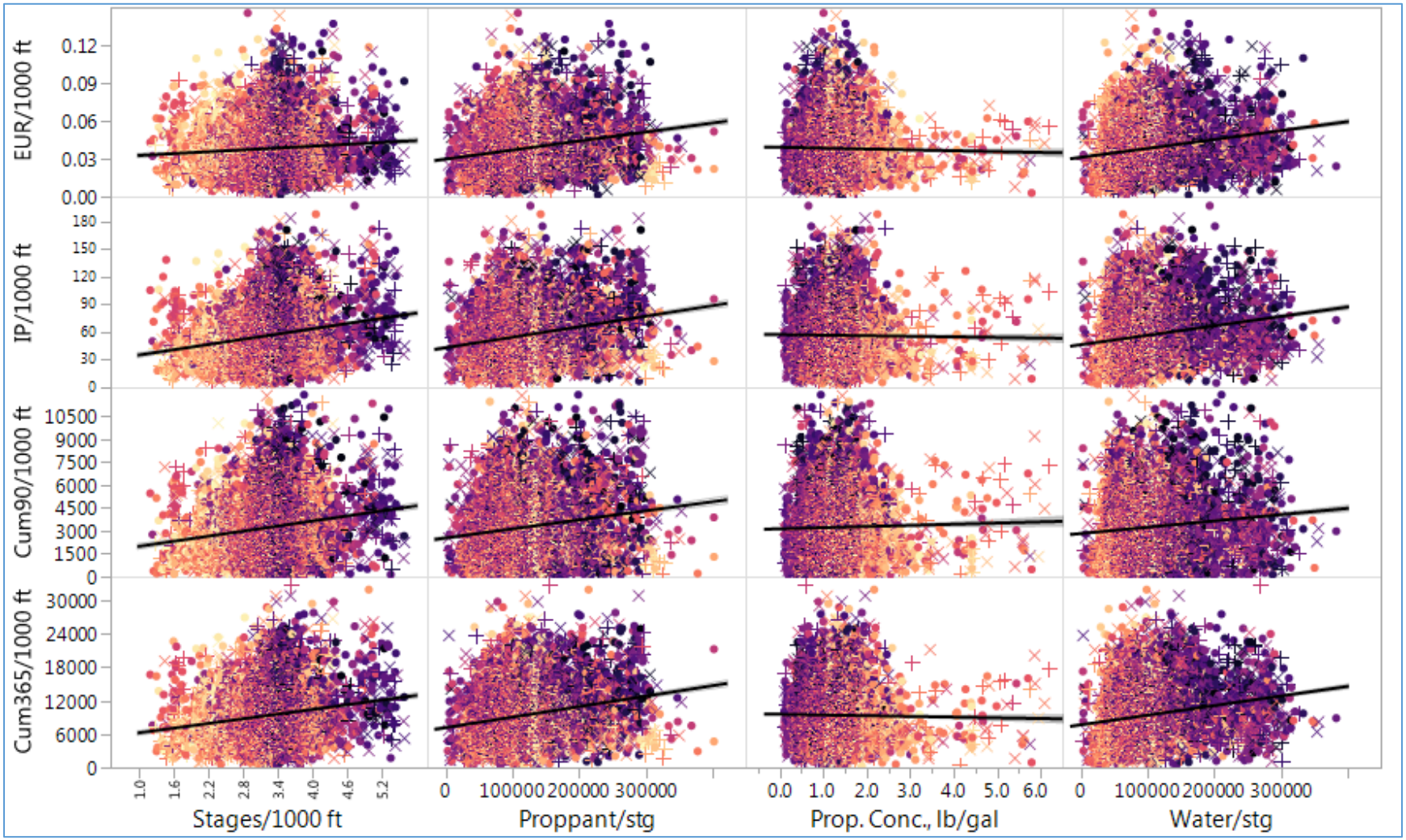


Fig. 5.7—Production metrics normalized with respect to Lateral Length vs. Stages/1,000 ft of Lateral Length, Proppant/stage, Proppant Concentration, and Water/stage. Darker points mean more recent Completion Date. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

The same behavior is generally true for production metrics normalized with respect to Stages (**Fig. 5.8**). However, these normalized metrics decrease with reducing fracture spacing, which agrees with Al-Alwani et al. (2015). This is expected since the stimulated reservoir volume (SRV) per stage decreases with decreasing fracture spacing. Thus less oil would be available for production per stage. **Fig. 5.9** provides a generalized plot for the four responses and their normalized values based on Stages and Lateral Length. The general trends of the 12 production metrics are plotted vs. Quarter, Stages, Stages/1,000 ft, Lateral Length, and Proppant Concentration are described below.

- Quarter: All the nine metrics related to Cum365, Cum90, and IP increase with time (Quarter). This might be attributed to the increase in completion productive efficiency with time. This trend is less pronounced for the metrics normalized based on Stages though. The situation is different, however, for the three EUR metrics. The recent developments in completion were successful in accelerating production, which is translated into higher Cum365, Cum90, and IP metrics. However, the SRV are depleted more quickly and the production decline rate increases. Therefore, EUR does not vary much. This agrees with the findings in Chapters 3 and 4.
- Fracture Stages: Stages-normalized metrics decrease with Stages increase. However, the non-normalized as well as Lateral Length-normalized production metrics increase with Stages. This implies that production efficiency decreases with increasing the Stages. This could be because the positive effect of adding more production pathways by extra Stages exceeds the negative effect of decreasing production due to the decrease in SRV per Stage. A fivefold increase in Stages

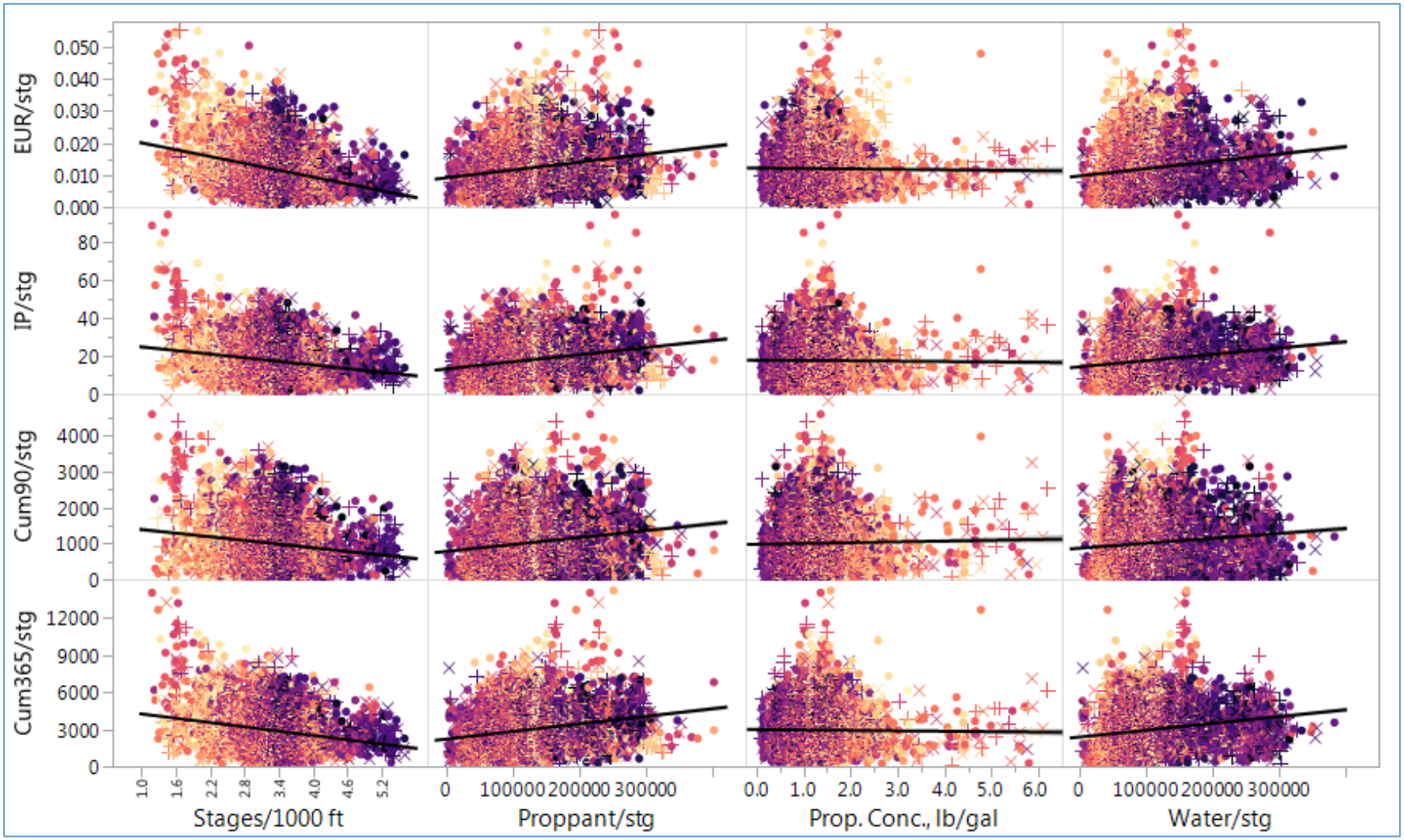


Fig. 5.8—Production metrics normalized with respect to Stages vs. Stages/1,000 ft of Lateral Length, Proppant/stage, Proppant Concentration, and Water/stage. Darker points mean more recent Completion Date. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

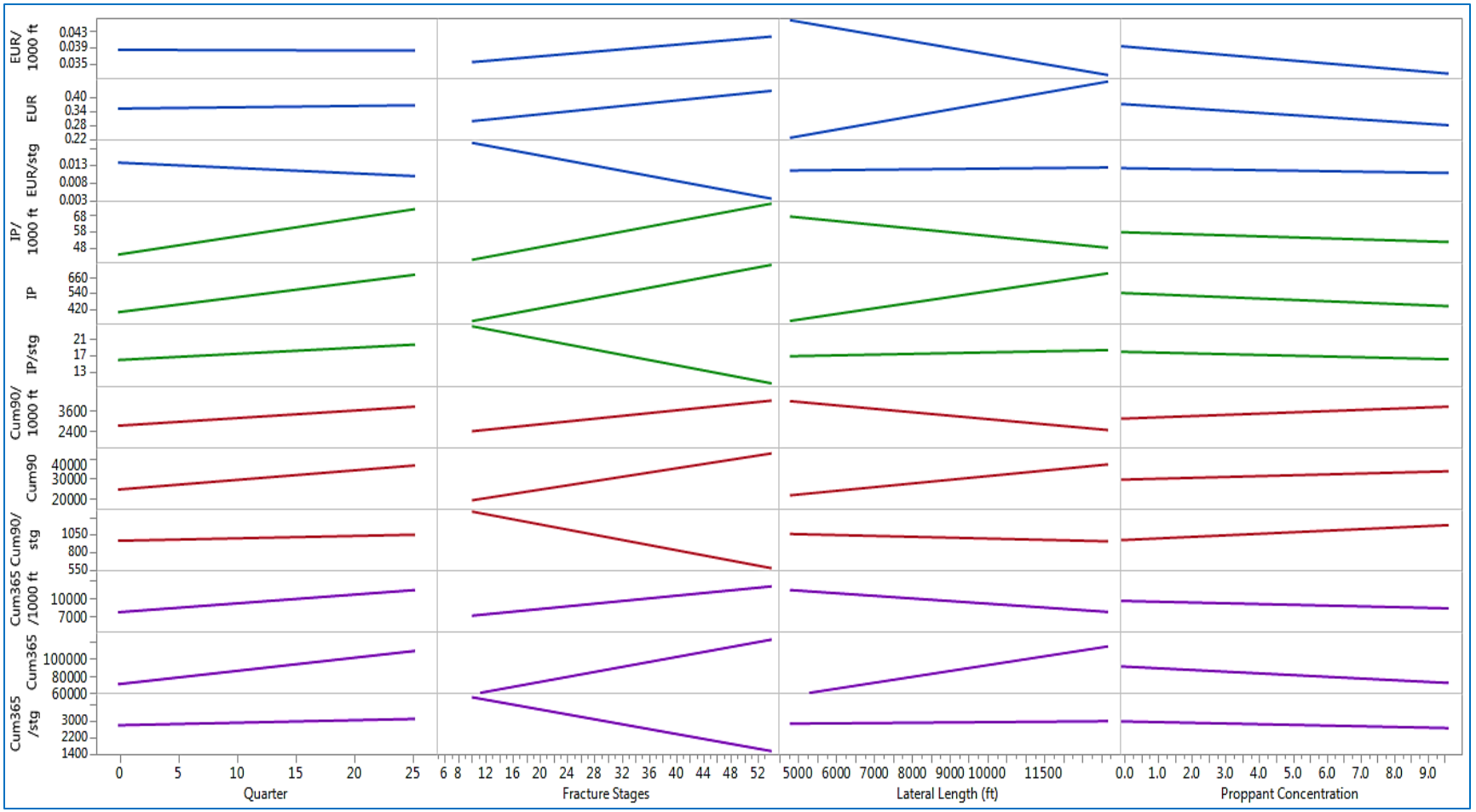


Fig. 5.9—Twelve production metrics vs. three predictors and Proppant concentration

from 10 to 50 translates into an increase of 2.2, 1.7, 2.0, and 1.4 times for IP, Cum90, Cum365, and EUR, respectively. This agrees with the previous finding that the effect of completion optimization has a larger impact on accelerating recovery than increasing its ultimate value.

- Lateral Length: Non-normalized metric increase with Lateral Length. However, doubling the Lateral Length does not necessarily mean that production doubles. This might be because of the decreased production efficiency when Lateral Length increases. Production efficiency reduction could be due to reduced efficiency to transport proppants to the toe of the lateral. This means that production per 1,000 ft is not constant along the lateral. It decreases with it. In addition, the farthest fractures along the lateral will have to have higher pressure for the oil produced to move longer distances. This interpretation is also manifested in the decrease of Lateral Length-normalized metrics with Lateral Length. Stages-normalized metrics are not affected by Lateral Length as expected.
- Proppant Concentration: The general trends of the 12 production metrics vs. Proppant Concentration are weak. The top performing wells have 1 – 2 lb/gal Proppant Concentration.

For Water and Proppant along with their normalized values based on Stages and Lateral Length, the 12-production metrics increase with all of them. Thus they were excluded from the graph for the purpose of clarity.

5.3.2 Supervised Learning (SL) Models and Their Evaluation

The best number of nodes that minimizes the errors and improves the fit of ANN is 26 for IP, EUR, and Cum90, and 30 for Cum365. These numbers minimize RMSE and MAE, and maximize R^2 for both training and validation sets. For RT, the best numbers of splits are 48, 15, 26, and 36 for IP, EUR, Cum90, and Cum365, respectively. These numbers minimize RMSE and maximize R^2 for both training and validation sets. For BF, the best numbers of trees that maximize validation R^2 are 27, 153, 128, and 66 for IP, EUR, Cum90, and Cum365, respectively. For BT, the best numbers of trees that maximize validation R^2 are 93, 121, 127, and 85 for IP, EUR, Cum90, and Cum365, respectively. For KNN, the best numbers of neighbors are 9, 18, 9, and 10 for IP, EUR, Cum90, and Cum365, respectively. These numbers minimize validation RMSE (for minimum training RMSE, these numbers are 8, 21, 5, and 12, respectively).

For LR, Box-Cox transformation was successful to stabilize the variance of IP and EUR. However, non-constant variance was still present for Cum90 and Cum365. This suggests that some variables might be missing. The absence of actual reservoir properties - which are expected to be much more influential than the completion and drilling variables (Chapter 4) - is believed to be the reason for that. This belief is supported by the fact that the location of the wells (Latitude and Longitude) is the most important variable (as shall be seen later).

For ENS, although median predictions are less affected by outliers of predicted values, they were found less accurate than the simple average. Arranged from least to most accurate, the ensembles are median-based, average-based, weighted average-based, and

best fit-based. Having the highest predictive power, best-fit ensembles would only be considered in further analysis.

Table 5.2 to **Table 5.5** provide a comparison between the various evaluation measures for the four responses. The measures of the training set for all 12 SL techniques of the four responses are very close to their corresponding values of the validation set. This indicates that overfitting probably was not a problem. For all responses, ENS always provides the best measures for both training and validation sets. **Fig. 5.10** shows how the relative errors (RE) of the four responses are distributed. **Fig. 5.11** provides a more clear distribution of REs using a log scale for their absolute values (AREs). The AREs cover a large spectrum of errors (1,000s %). However, as **Table 5.6** shows, more than 90% of the AREs are below 50% (REs lie within ± 50%) except for Cum90, which has 83.92% of the AREs below 50%. Color-coding based on the logarithm of the response (Fig. 5.11) reveals that all the large REs occur for the poor producing wells. This is expected since a small deviation of the predicted values from the actual ones would lead to large REs for such wells. Table 5.7 also demonstrates that for all four production metrics, ENS consistently provides the highest percentage of data points that lie within any interval of ARE. This agrees with the previous conclusion that ENS has the highest predictive power.

Accuracy metric	Absolute								Relative					
	R^2		RMSE		MAE		MdAE		MARE		MdARE		RSE	
	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld
ENS	**0.744**	**0.761**	**131.9**	**134.6**	**96.7**	**98.0**	**72.5**	**71.8**	**23.9**	**24.3**	**15.0**	**14.8**	**0.256**	**0.239**
BT	0.686	0.711	145.9	147.9	109.8	108.8	86.7	83.2	28.0	27.4	17.4	16.6	0.314	0.289
BF	0.687	0.701	145.7	150.5	108.2	110.5	81.3	82.2	28.0	28.6	16.7	16.5	0.313	0.299
ANN	0.505	0.534	183.3	187.8	140.5	140.3	113.6	105.5	36.4	36.2	23.0	21.7	0.495	0.466
KNN	0.499	0.525	184.5	189.6	141.5	140.7	115.7	108.9	37.6	38.3	22.7	21.5	0.501	0.475
RT	0.464	0.483	190.8	197.9	145.5	147.4	117.2	112.3	37.7	37.5	23.5	22.2	0.536	0.517
LASSO	0.377	0.388	205.7	215.2	159.2	164.5	129.7	131.8	41.9	43.4	26.0	26.4	0.623	0.612
RR	0.378	0.387	205.5	215.5	159.0	165.0	131.1	131.0	41.8	43.5	26.0	26.3	0.622	0.613
EN	0.378	0.387	205.5	215.5	159.0	165.0	131.1	131.0	41.8	43.5	26.0	26.3	0.622	0.613
A-LASSO	0.358	0.386	208.8	215.6	161.1	164.6	128.7	130.5	42.7	43.2	25.8	26.2	0.642	0.614
A-EN	0.358	0.385	208.9	215.7	161.1	164.7	128.9	129.6	42.7	43.3	25.8	26.2	0.642	0.615
LR	0.366	0.383	207.5	216.2	157.8	161.0	122.8	118.9	39.3	40.1	25.4	25.1	0.634	0.617

Table 5.2—Accuracy metrics for IP arranged ascendingly according to validation RMSE.

Accuracy metric	Absolute								Relative					
	R^2		RMSE		MAE		MdAE		MARE		MdARE		RSE	
	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld
ENS	**0.761**	**0.746**	**0.088**	**0.090**	**0.062**	**0.064**	**0.044**	**0.045**	**22.1**	**23.0**	**13.8**	**14.2**	**0.239**	**0.254**
BT	0.696	0.682	0.099	0.101	0.072	0.073	0.051	0.053	26.2	27.2	15.7	16.3	0.304	0.318
BF	0.665	0.649	0.104	0.106	0.075	0.078	0.053	0.057	27.6	29.6	16.5	17.2	0.335	0.351
ANN	0.444	0.440	0.134	0.134	0.099	0.100	0.077	0.076	36.6	38.4	23.1	23.1	0.556	0.560
KNN	0.391	0.393	0.140	0.140	0.105	0.105	0.082	0.083	39.6	41.3	24.3	24.4	0.609	0.607
LASSO	0.328	0.324	0.147	0.147	0.111	0.112	0.089	0.088	41.4	43.3	26.3	26.7	0.672	0.676
EN	0.327	0.324	0.147	0.148	0.111	0.112	0.089	0.089	41.4	43.4	26.3	26.8	0.673	0.676
A-LASSO	0.314	0.323	0.149	0.148	0.113	0.112	0.091	0.088	42.2	43.5	27.0	26.7	0.686	0.677
RR	0.328	0.323	0.147	0.148	0.111	0.112	0.090	0.088	41.4	43.4	26.3	26.7	0.672	0.677
A-EN	0.328	0.323	0.147	0.148	0.111	0.112	0.090	0.088	41.4	43.4	26.3	26.7	0.672	0.677
LR	0.315	0.312	0.148	0.149	0.109	0.110	0.083	0.083	37.8	39.6	26.1	26.2	0.685	0.688
RT	0.329	0.312	0.147	0.149	0.111	0.113	0.087	0.088	41.5	44.3	25.9	26.2	0.671	0.688

Table 5.3—Accuracy metrics for EUR arranged ascendingly according to validation RMSE.

Accuracy metric	Absolute								Relative					
	R^2		RMSE		MAE		MdAE		MARE		MdARE		RSE	
	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld
ENS	**0.793**	**0.784**	**19293**	**20085**	**13686**	**14169**	**9953**	**10082**	**21.4**	**20.9**	**12.1**	**12.4**	**0.207**	**0.216**
BT	0.755	0.754	20985	21422	15050	15441	10994	11222	24.7	23.8	13.1	13.8	0.245	0.246
BF	0.709	0.714	22906	23101	16447	16760	11874	12149	27.9	26.8	14.3	14.6	0.291	0.286
ANN	0.480	0.519	30584	29979	22891	22568	17473	17481	38.4	35.7	21.3	20.6	0.520	0.481
KNN	0.466	0.484	31011	31024	23325	23418	18178	18199	39.9	38.8	21.5	21.3	0.534	0.516
RT	0.417	0.445	32401	32187	24147	24478	18393	19292	40.9	39.5	22.5	22.9	0.583	0.555
A-LASSO	0.313	0.356	35164	34683	26851	26803	21434	21424	46.2	43.6	25.6	25.0	0.687	0.644
LASSO	0.329	0.355	34747	34690	26469	26787	20918	21129	45.4	43.6	25.1	25.4	0.671	0.645
A-EN	0.313	0.355	35173	34695	26860	26813	21417	21608	46.3	43.6	25.6	25.0	0.687	0.645
RR	0.331	0.354	34714	34727	26448	26843	20941	20977	45.3	43.6	25.2	25.5	0.669	0.646
EN	0.331	0.354	34714	34727	26448	26843	20941	20977	45.3	43.6	25.2	25.5	0.669	0.646
LR	0.318	0.346	35040	34931	26349	26624	20501	20470	43.2	41.3	24.8	24.8	0.682	0.654

Table 5.4—Accuracy metrics for Cum90 arranged ascendingly according to validation RMSE.

Accuracy metric	Absolute								Relative					
	R^2		RMSE		MAE		MdAE		MARE		MdARE		RSE	
	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld	Trn	Vld
ENS	**0.744**	**0.761**	**9292**	**9234**	**6649**	**6554**	**4924**	**4459**	**45.7**	**42.9**	**18.0**	**17.4**	**0.256**	**0.239**
BT	0.685	0.700	10305	10349	7597	7723	5839	5895	58.4	58.8	20.3	20.8	0.315	0.300
BF	0.664	0.680	10651	10687	7827	7957	5829	5881	62.2	62.6	20.2	22.1	0.336	0.320
KNN	0.406	0.428	14155	14293	10675	10826	8225	8362	83.5	90.5	28.8	28.3	0.594	0.572
ANN	0.357	0.363	14732	15076	11135	11529	8492	8737	90.1	96.0	29.0	30.0	0.643	0.637
RT	0.317	0.329	15185	15476	11526	11902	8855	9433	91.9	98.1	30.5	32.2	0.683	0.671
LASSO	0.230	0.230	16119	16580	12350	12840	9813	10035	99.7	106.8	32.7	34.6	0.770	0.770
RR	0.231	0.229	16110	16584	12351	12849	9742	9895	99.3	106.5	32.8	34.3	0.769	0.771
EN	0.231	0.229	16110	16584	12351	12849	9742	9895	99.3	106.5	32.8	34.3	0.769	0.771
A-EN	0.231	0.229	16110	16584	12351	12849	9742	9895	99.3	106.5	32.8	34.3	0.769	0.771
A-LASSO	0.231	0.229	16110	16584	12351	12849	9742	9895	99.3	106.5	32.8	34.3	0.769	0.771
LR	0.198	0.206	16456	16837	12457	12886	9611	10260	93.2	98.9	33.6	36.0	0.802	0.794

Table 5.5—Accuracy metrics for Cum365 arranged ascendingly according to validation RMSE.

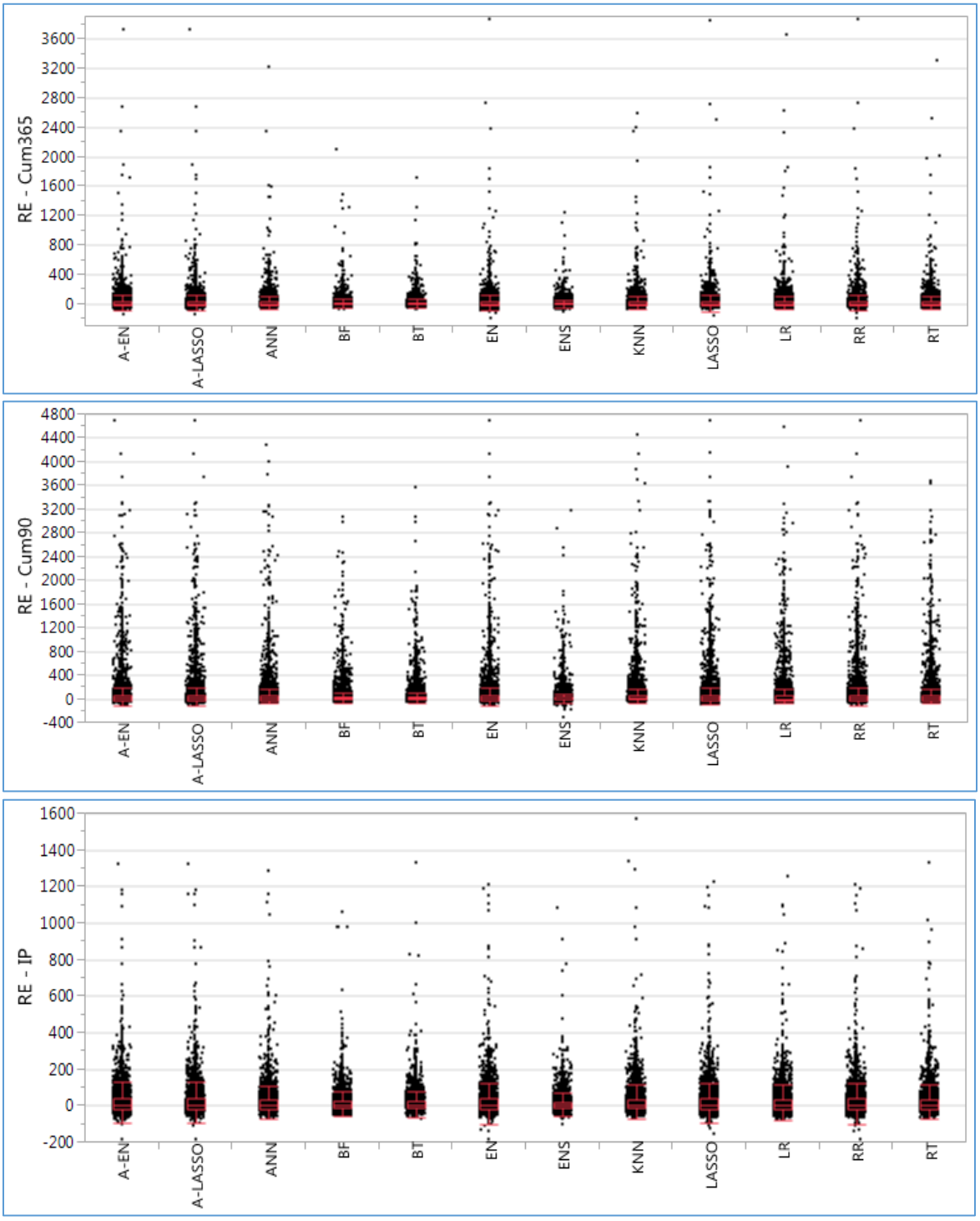


Fig. 5.10—Boxplots showing the distributions of REs for the four production responses.

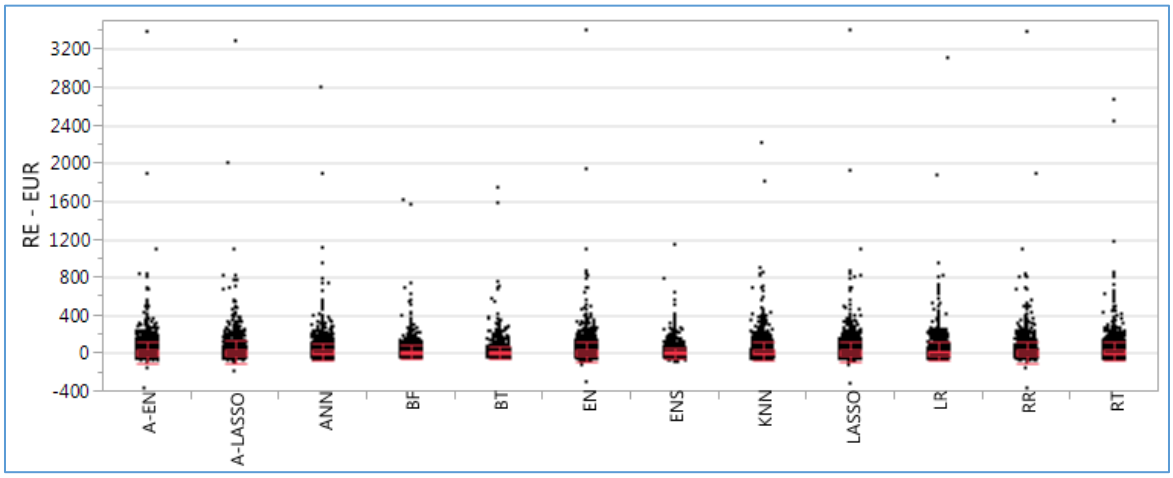


Fig. 5.10 Continued—Boxplots showing the distributions of REs for the four production responses.

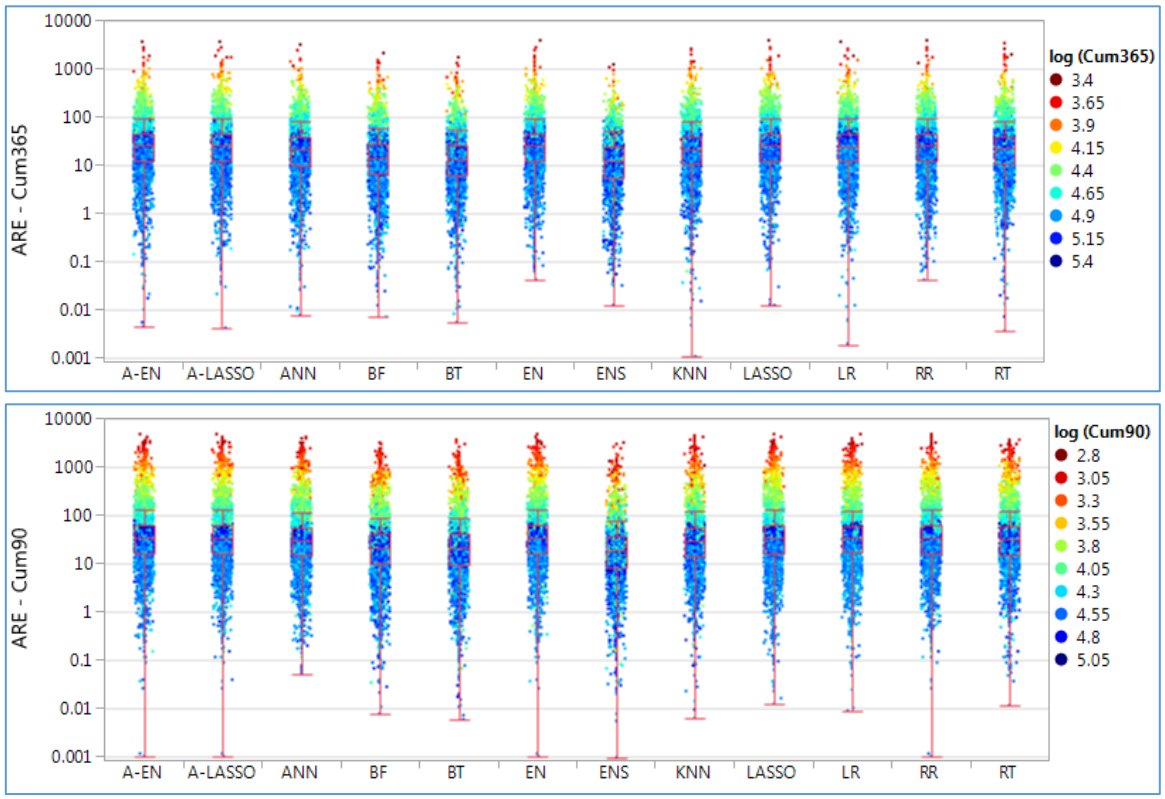


Fig. 5.11—Boxplots showing the distributions of AREs for the four production responses. Log scale used for clarity.

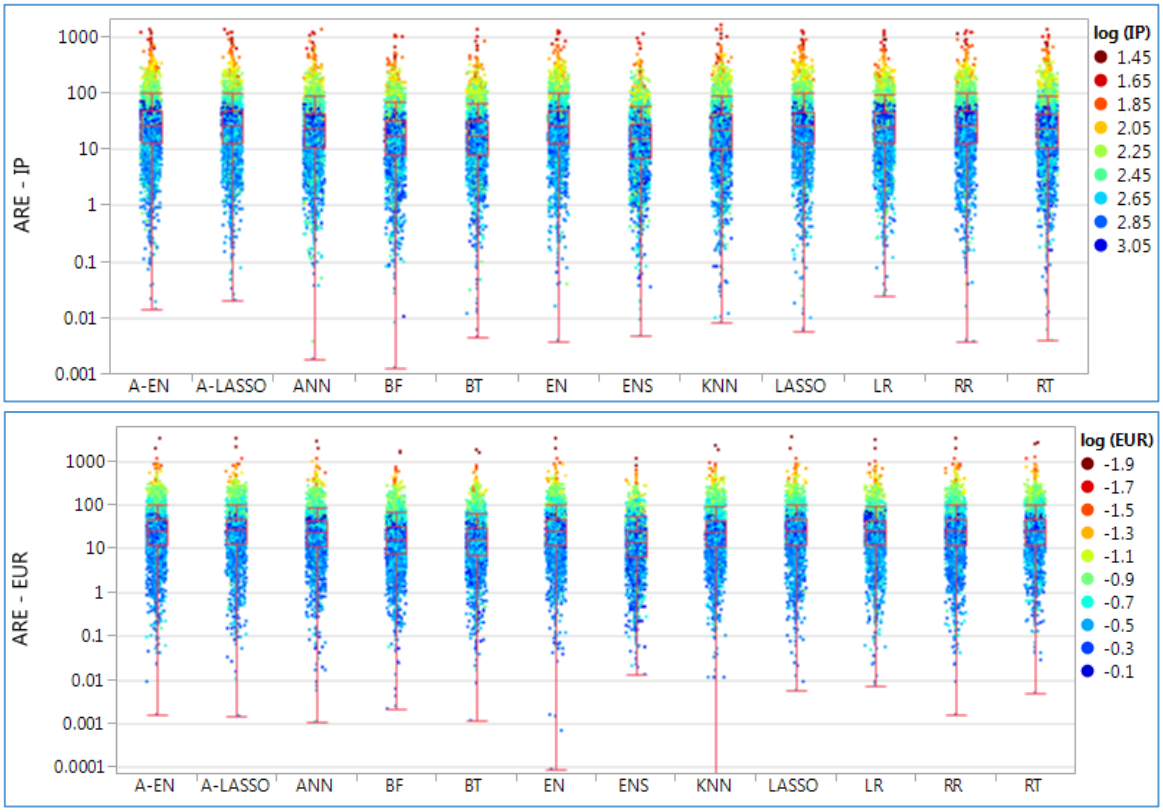


Fig. 5.11 Continued—Boxplots showing the distributions of AREs for the four production responses. Log scale used for clarity.

Response	IP						Response	EUR					
ARE	≤ 10%	≤ 20%	≤ 30%	≤ 50%	≤ 70%	≤ 100%	ARE	≤ 10%	≤ 20%	≤ 30%	≤ 50%	≤ 70%	≤ 100%
EN	19.41	39.37	55.25	77.79	85.29	90.67	A-EN	19.01	38.62	54.70	76.23	85.90	91.76
RR	19.41	39.37	55.25	77.79	85.29	90.67	RR	19.01	38.62	54.70	76.23	85.90	91.76
A-LASSO	19.69	40.26	55.52	77.72	85.49	90.87	EN	19.62	38.28	55.25	76.57	85.90	91.42
A-EN	19.75	40.40	55.59	77.72	85.56	90.87	LASSO	19.62	38.22	55.45	76.57	86.04	91.49
LASSO	19.82	39.37	55.99	77.72	85.29	90.74	LR	20.23	38.56	55.79	80.11	89.03	93.39
LR	20.78	40.94	58.17	79.77	87.13	92.23	A-LASSO	19.28	38.35	55.86	76.43	86.10	91.55
RT	25.34	45.50	62.47	79.29	87.53	92.71	RT	19.48	40.12	55.99	77.18	85.22	90.05
ANN	24.59	46.80	64.44	82.08	88.49	93.46	KNN	22.55	42.51	58.86	78.20	86.38	91.35
KNN	26.43	46.39	64.58	81.68	88.08	92.37	ANN	22.75	44.28	61.58	80.79	88.28	92.57
BF	31.68	58.65	73.57	86.51	92.30	95.78	BF	30.31	55.52	72.07	85.97	91.49	95.84
BT	32.43	58.24	74.25	87.74	93.60	96.05	BT	33.65	58.04	73.64	88.15	93.19	96.73
ENS	36.99	63.49	78.41	90.19	94.14	96.59	ENS	38.69	63.28	78.27	91.08	95.50	98.02

Response	Cum90						Response	Cum365					
ARE	≤ 10%	≤ 20%	≤ 30%	≤ 50%	≤ 70%	≤ 100%	ARE	≤ 10%	≤ 20%	≤ 30%	≤ 50%	≤ 70%	≤ 100%
LR	14.65	28.68	42.37	66.28	77.86	82.08	LASSO	20.71	40.19	57.56	78.34	86.44	91.01
A-EN	15.33	29.97	43.53	66.01	74.93	80.04	EN	20.10	40.19	57.77	78.34	86.44	91.01
A-LASSO	15.33	29.97	43.53	66.01	74.93	80.04	RR	20.10	40.19	57.77	78.34	86.44	91.01
EN	15.33	29.97	43.53	66.01	74.93	80.04	A-LASSO	21.59	40.74	58.31	77.93	86.10	90.94
RR	15.33	29.97	43.53	66.01	74.93	80.04	A-EN	21.87	40.80	58.38	77.93	86.04	90.94
LASSO	16.08	30.11	43.60	66.01	75.20	80.25	LR	20.23	40.94	58.38	80.52	88.08	91.89
RT	16.89	32.49	47.55	69.48	76.77	82.02	RT	23.43	44.07	62.87	81.95	87.87	92.57
ANN	16.83	33.92	50.00	70.91	78.13	82.97	KNN	24.59	47.14	64.85	82.63	88.15	92.03
KNN	18.80	35.97	52.79	71.46	79.02	83.58	ANN	25.89	48.57	66.62	83.92	89.17	93.26
BF	25.75	45.98	64.65	78.41	84.33	88.22	BF	36.38	61.44	76.98	88.96	93.19	96.12
BT	25.95	48.50	65.53	79.90	84.88	88.90	BT	37.94	65.05	81.13	90.12	93.60	97.00
ENS	30.86	55.04	69.69	83.92	89.03	92.85	ENS	41.69	69.01	83.51	92.17	95.71	98.02

Table 5.6—Distribution of validation AREs arranged ascendingly according to 30% ARE.

The next best models after ENS are BT and BF. This is expected because these BT and BF are ensemble models themselves. Ensemble is powerful since it enhances the robustness and accuracy of predictions. The six regression techniques are the worst in terms of their predictive performance.

Being the best model, ENS was then tested using the test set to provide an unbiased measure of its performance. **Table 5.7** shows the distribution of AREs for ENS. AREs of training and validation sets are also included for comparison. It is obvious that the predictive performance of ENS is consistently lower for the test set. This is also supported by the predictive performance metrics provided in **Table 5.8**. **Fig. 5.12** gives a boxplot distribution for the REs and AREs for the three sets. It is apparent that the test set has higher percentage of larger REs. In addition to the fact that most of the large REs occur for the poor producing wells, the absence of the overwhelmingly important reservoir characteristics, and the presence of unavoidable error in the available dataset could be the reason for that deterioration in the predictive performance of the model. **Fig. 5.13** shows the actual-by-predicted plots for the four responses along with the distribution of the errors. It shows that the test points are well distributed around the unit-slope line. Therefore, the predictive performance of ENS was considered acceptable.

Resp.	Set	≤ 10%	≤ 20%	≤ 30%	≤ 50%	≤ 70%	≤ 100%
IP	Trn	35.20	62.28	77.44	90.53	94.68	97.24
	Vld	36.99	63.49	78.41	90.19	94.14	96.59
	Tst	24.05	46.72	62.96	82.86	90.37	94.32
EUR	Trn	38.18	64.69	80.20	91.74	95.33	97.67
	Vld	38.69	63.28	78.27	91.08	95.50	98.02
	Tst	24.30	45.68	61.98	82.02	90.52	94.86
Cum90	Trn	30.12	54.22	69.71	83.55	88.96	92.09
	Vld	30.86	55.04	69.69	83.92	89.03	92.85
	Tst	19.41	38.47	53.53	75.21	83.21	88.10
Cum365	Trn	42.70	69.23	83.07	92.68	95.55	97.59
	Vld	41.69	69.01	83.51	92.17	95.71	98.02
	Tst	25.63	48.94	67.56	85.28	91.60	95.11

Table 5.7—Distribution of the test (Tst) set AREs for ENS. AREs of training (Trn) and validation (Vld) sets are included for comparison.

Resp.	Set	R^2	RMSE	MAE	MdAE	MARE	MdARE	RSE
IP	Trn	0.7439	131.88	96.65	72.52	23.92	14.99	0.2561
	Vld	0.7609	134.56	97.99	71.80	24.31	14.77	0.2391
	Tst	0.5136	186.24	139.46	105.57	34.92	21.68	0.4864
EUR	Trn	0.7610	0.0876	0.0623	0.0442	22.07	13.81	0.2390
	Vld	0.7462	0.0904	0.0640	0.0451	22.97	14.15	0.2538
	Tst	0.3877	0.1402	0.1006	0.0720	36.01	22.16	0.6123
Cum90	Trn	0.7442	9292	6649	4924	45.74	18.00	0.2558
	Vld	0.7611	9234	6554	4459	42.94	17.35	0.2389
	Tst	0.4507	13644	10081	7394	69.57	27.72	0.5493
Cum365	Trn	0.7932	19293	13686	9953	21.40	12.13	0.2068
	Vld	0.7839	20085	14169	10082	20.95	12.44	0.2161
	Tst	0.5039	29784	22061	16255	34.52	20.41	0.4961

Table 5.8—Predictive performance metrics of the testing (Tst) set for ENS. Metrics of Training (Trn) and validation (Vld) sets are included for comparison.

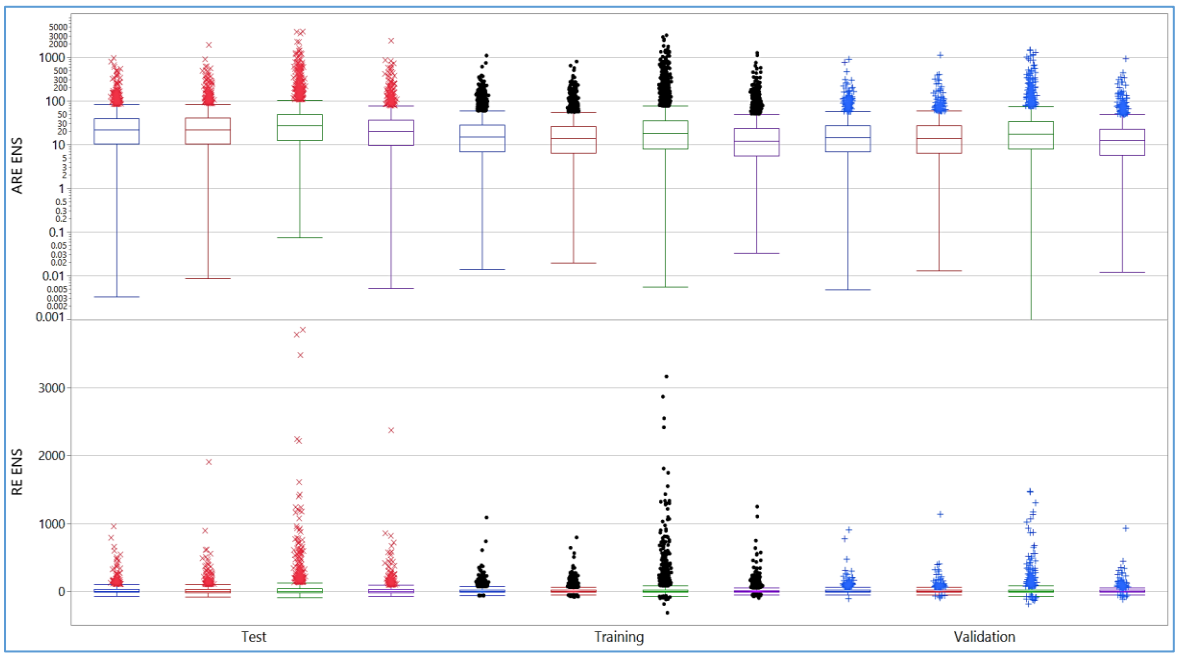


Fig. 5.12—Boxplots of REs, and AREs for the four responses.

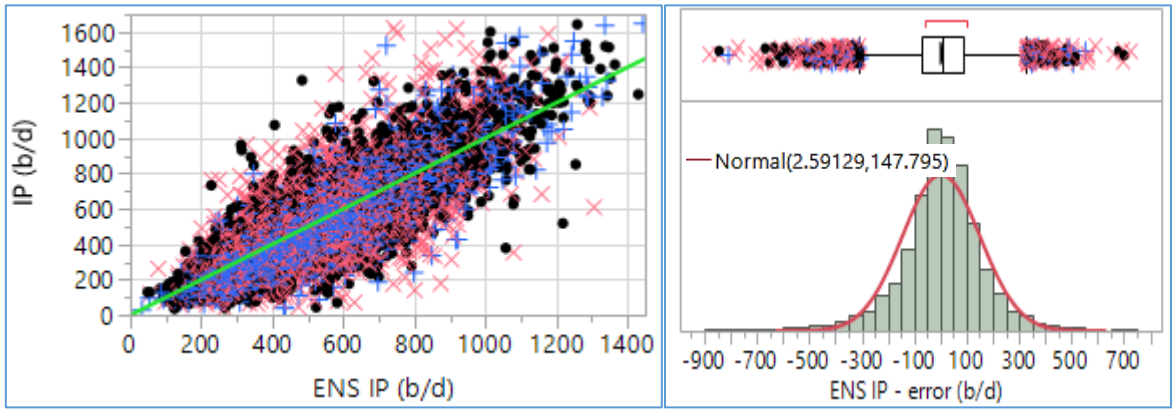


Fig. 5.13—Actual-by-predicted plots for the four responses along with the distribution of the errors. Training, validation, and test sets shown in black dots, red pluses, and blue crosses, respectively.

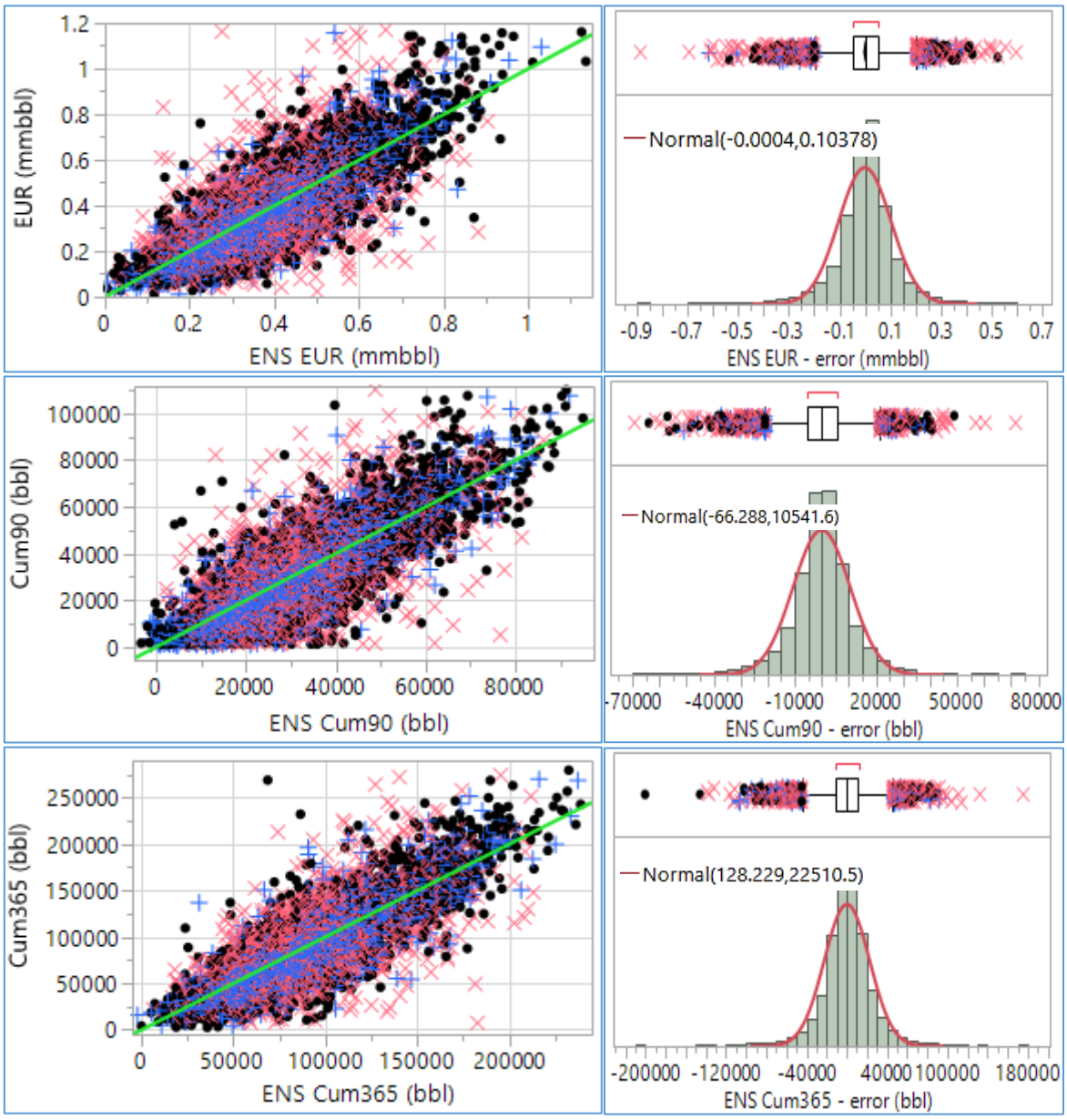


Fig. 5.13 Continued—Actual-by-predicted plots for the four responses along with the distribution of the errors. Training, validation, and test sets shown in black dots, red pluses, and blue crosses, respectively.

5.3.3 Importance of Variables

Fig. 5.14 shows how different variables rank in terms of their importance for the four responses of the different SL techniques. For the six regression techniques, because there were many important squared terms and interactions, only the main effects were shown in this figure.

Latitude and Longitude are dominantly important for all the four responses as evaluated by various techniques. They are also included in the most important squared terms and interactions in the regression techniques. The reason for such dominance could be due to the fact that reservoir characteristics are much more important than design variables related to drilling and hydraulic fracturing. Since such characteristics are not available for analysis, well location can be considered as their indicator because the reservoir sweet spots tend to be allocated to certain ranges of Latitudes and Longitudes. This agrees with Izadi et al. (2013) who considered well location as a proxy for reservoir quality. This is also consistent with the maps of thermal maturity, pore pressure gradient, and hydrocarbon pore volume of Kuhn et al. (2010), Wescott (2016), and Hamlin et al. (2017), respectively. These maps have contour patterns that are very similar to the Cum365 map (Fig. 5.4).

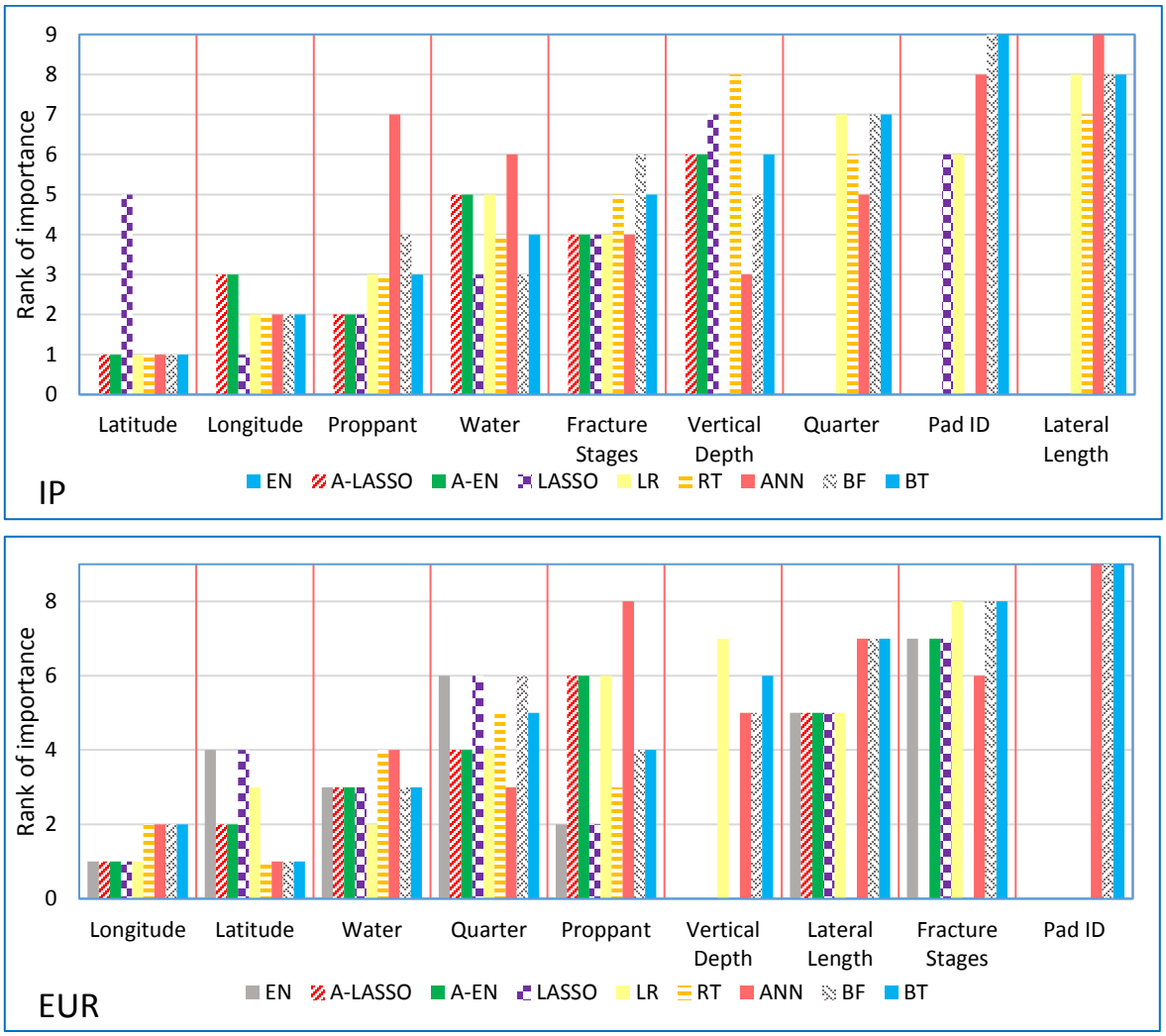


Fig. 5.14—Rank of importance of the nine predictors for all four responses with nine SL techniques.

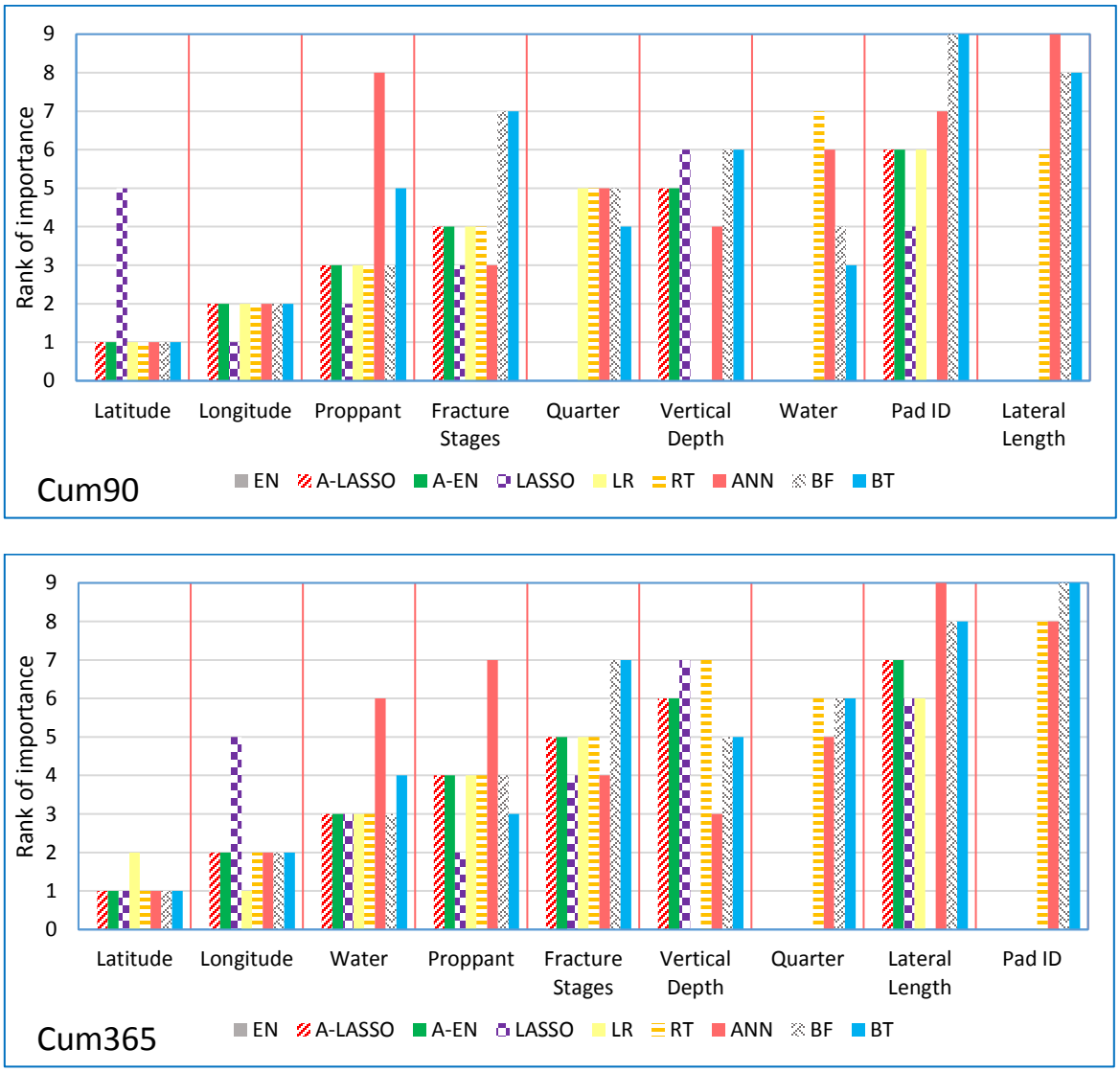


Fig. 5.14 Continued—Rank of importance of the nine predictors for all four responses with nine SL techniques.

Lateral Length is one of the least important variables. This could be mainly because of its very small variability. It did not change much (Fig. 5.6). Therefore, the SL techniques were not able to capture its effect on the production metrics. This would have been different if there had been more variation in Lateral Length. Another minor reason could be the inefficiency associated with longer laterals (Fig. 5.9). This agrees with Jain et

al. (2013) who found that some fracture clusters do not contribute to production, due to the non-uniform distribution of proppant and fluids.

Fig. 5.15 shows the bivariate relationships between the four responses and the nine predictors. Although the bivariate relationships are weak (no R^2 exceeds 0.1), the figure displays the general trend of the effects of the predictors on the responses. In general, the four production metrics are improved with increasing Longitude, Vertical Depth, Lateral Length, Water, Proppant, Fracture Stages, and time (Quarter), and if a well is on a pad (the weakest relationship). The only predictor that has a negative effect is the Latitude. Despite being weak, the variable that has the largest R^2 is Vertical Depth. However, this variable is ranked 6^{th} on average for all four responses (Fig. 5.14). This conflict could be resolved by observing that Vertical Depth is also highly correlated with Latitude (Table 5.1). Thus the effect of Vertical Depth is confounded with that of Latitude, the most important variable for all four responses on average. Accordingly Vertical Depth is believed to be more important than evaluated by the different SL methods, which agrees with the findings of Wang and Chen (2016a). This is also clearer if the map of **Fig. 5.16** is compared with Fig. 5.4. In fact, no other completion variable has a strong match between its map (like Fig. 5.16) and any production response map (like Fig. 5.4) as does Vertical Depth. This makes since pressure, which is the main driving force of production, is a function of vertical depth.

Fig. 5.15—Effects of the nine predictors on the four responses.

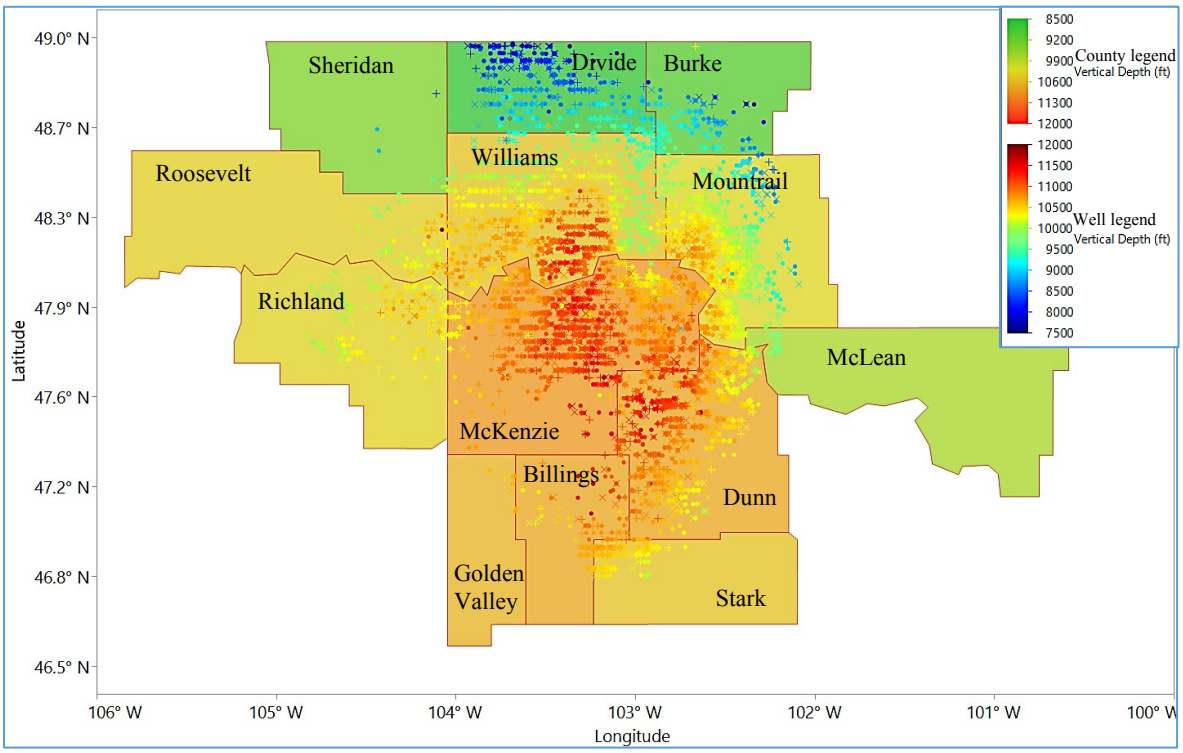


Fig. 5.16—County map for Bakken play wells color-coded based on Vertical Depth. Training, validation, and test sets shown in dots, pluses, and crosses, respectively.

5.4 Conclusions

The data analytics framework provided here could help produce statistically-based conclusions and data-driven facts, which ensures the objectivity of the decision-making process. Harnessing the evolutionary potentials of data analytics, this chapter applies 12 SL techniques to mine data-driven insights from the available big data and translate them into better understanding of Bakken shale play and improving its production performance, and to identify key factors that distinguish good wells from poor-performing ones. The main findings and conclusions of the study are as follows:

1. The optimization of completion design and well architecture can accelerate oil recovery. The effect of such a process on improving the RF is limited. Exploring EOR is important to improve the RF.
2. Ensembles model outperforms the other 11 SL models. It can predict can predict the four responses in the test set with an average of 61.5% and 81.4% of the predictions less than 30% and 50%, respectively, of absolute relative error. Such a model can be used to optimize the location of future wells, their architecture, and hydraulic fracture design, and predict their production.
3. Reservoir quality is much more important than completion quality for reservoir performance. To improve oil recovery, searching for sweet spots is more recommended that focusing on completion optimization.

6. SUMMARY AND RECOMMENDATIONS

In this dissertation, the main goal was to apply a systematic DoE-based workflow to improve shale oil recovery using simulation. Four major recovery schemes were studied for Eagle Ford shale. The four schemes are primary production, waterflooding, continuous miscible gas flooding, and miscible gas huff 'n' puff. This work is the first to systematically and statistically compare the performance of these four methods in Eagle Ford shale. Sequential DoE was applied to search for the maximum RF region in the multidimensional variable space. An innovative injection pattern that relies on alternating injection and producing fractures along the same lateral was considered. Here are the main findings, conclusions, and recommendations for Eagle Ford:

1. Fracture design and lateral placement are recommended to be designed such that fracture wings extend to half the distance between the laterals. Fracture spacing should be around 20 - 40 ft and fracture half-length should be maximized as practically as possible. This design is optimum for primary production as well as potential subsequent recovery schemes. No infill wells or re-fracturing is needed.
2. Being the two most influential design variables on RF, fracture spacing and fracture half-length should be better estimated in order to reduce the uncertainty in forecasting oil production.
3. PP optimization is not enough to improve the RF from Eagle Ford shale. Exploring other recovery methods is a necessity.
4. Continuous miscible gas flooding with injectors and producers along the same lateral was evaluated to have the highest potential to maximize RF.

5. Developing alternating injection/production fractures along the same lateral might be the next breakthrough to boost the RF from shale plays.

Because of simulation assumptions, data analytics was applied as well with the main objective of validating simulation results. Another case that involved primary production in Bakken was studied using DoE and simulation. Data analytics supported the findings in the simulation results. In addition, a model was developed that can reasonably predict future well production performance based on its potential location and completion design. The main findings and conclusions are:

1. Reservoir quality is more important than completion quality.
2. Completion and well design optimization can accelerate oil recovery. However, its effect on ultimate RF is limited by physical constraints. Thus it is recommended that research should focus more on finding sweet reservoir spots and EOR.
3. The best data analytics model (ensembles) has a practically acceptable prediction accuracy. It can be used to optimize the location of future wells, their architecture, and hydraulic fracture design, and predict their production.

The systematic framework presented here allowed for an efficient exploration of the multidimensional variable space to pinpoint the optimum spots where recovery could be maximized. This framework produces statistically-based conclusions and data-driven facts, which could improve the objectivity of the decision-making process.

The specific results presented in this dissertation are valid only for the input data used here. The conclusions should be valid for other sets of inputs that are not much

different than the data used in this work. The main purpose of this research is to illustrate the use of Design of Experiments in the oil and gas industry.

We hope that this work could help encourage petroleum researchers and engineers to incorporate statistics at the heart of industrial and academic research and problem-solving tools, and to learn and apply DoE regardless of the type of experiments that they conduct; physical experiments, simulation runs, or field trials.

REFERENCES

Agrawal, D., Dwivedi, S., Barrois, A., et al. 2015. Impact of Environmental Parameters on Forward Stratigraphic Modelling from Uncertainty Analysis; Lower Cretaceous, Abu Dhabi. SPE-175683-MS. http://dx.doi.org/10.2118/175683-MS.

Al-Alwani, M., Britt, L. K., and Dunn-Norman, S. 2015. Data Mining and Statistical Analysis of Completions in an Unconventional Play: The Canadian Montney Formation. SPE-178586-MS/URTEC-2154384. http://dx.doi.org/10.15530/URTEC-2015-2154384.

Alfarge, D., Wei, M., and Bai, B. 2017. IOR Methods in Unconventional Reservoirs of North America: Comprehensive Review. SPE-185640-MS. http://dx.doi.org/10.2118/185640-MS.

Alharthy, N. S. 2015. *Compositional Modeling of Multiphase Flow and Enhanced Oil Recovery Prospects in Liquid-Rich Unconventional Reservoirs*. PhD dissertation, Colorado School of Mines, Colorado (2015).

Al-Mudhafar, W.J. and Rao, D.N. 2015. Optimization of Gas Assisted Gravity Drainage (GAGD) Process in a Heterogeneous Sandstone Reservoir: Field-Scale Study. SPE-174579-MS. http://dx.doi.org/10.2118/174579-MS.

Amudo, C., Graf, T., Dandekar, R.R. et al. 2009. The Pains and Gains of Experimental Design and Response Surface Applications in Reservoir Simulation Studies. SPE-118709-MS. http://dx.doi.org/10.2118/118709-MS.

Anderson, M. J. and Whitcomb, P. J. 2017. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*, 2nd edition. Boca Raton, Florida: CRC Press.

Antony, J., Chou, T.Y., and Ghosh, S. 2003. Training for Design of Experiments. *Work Study* **52** (7): 341-346. http://dx.doi.org/10.1108/00438020310502642.

Apaydin, M. 2014. *Economic Profitability Of the Bakken, North Dakota Unconventional Oil Plays Based on a Typical Well Performance with Current Market Conditions*. MSc thesis, Michigan Technological University, Michigan (2014).

Austin, Peter C. and Steyerberg, Ewout W. 2015. The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, **68** (6), 627–636. http://dx.doi.org/10.1016/j.jclinepi.2014.12.014.

Bettonvil, B. and Kleijnen, J.P.C. 1997. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research* **96** (1): 180-194. http://dx.doi.org/10.1016/S0377-2217(96)00156-7.

Bingham, D., Sitter, R.R. and Tang, B. 2009. Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* **96** (1): 51-65. http://dx.doi.org/10.1093/biomet/asn057.

Booker, A.J., Dennis, J.E., Jr., Frank, P.D. et al. 1999: A rigorous framework for optimization of expensive functions by surrogates. *Struct. Optim.* **17** (1): 1–13. http://dx.doi.org/10.1007/BF01197708.

Box, G.E.P. 1979. Robustness in the strategy of scientific model building. In *Robustness in Statistics*, Launer, R.L. and Wilkinson, G.N., 201–236, Academic Press.

Box, G.E.P. 2001. Statistics for discovery. In *Journal of Applied Statistics* **28** (3 - 4): 285-299. http://dx.doi.org/10.1080/02664760120034036.

Box, G.E.P., Hunter, J.S., and Hunter, W.G. 2005. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, 2nd edition. NY: John Wiley & Sons.

Carson II, J. S. 2002. Model verification and validation. *Proc*., 2002 IEEE Winter Simulation Conference, 8-11 December, 52-58. http://dx.doi.org/10.1109/WSC.2002.1172868.

Castillo, L.M., Jablonowski, C.J., and Olson, J.E. 2010. Integrated Analysis to Optimize Hydraulic Fracturing Treatment Design in a Clastic and Carbonate Formation in Torunos Hydrocarbon Field, Barinas: Apure Basin, Southwest Venezuela. SPE-132511. http://dx.doi.org/10.2118/132511-MS.

Cinco-Ley, H., and Samaniego, F. 1981. Transient Pressure Analysis for Fractured Wells. *J Pet Technol* **33** (9): 1749–1766. SPE-7490-PA. http://dx.doi.org/10.2118/7490-PA.

Chapman, W.L., Welch, W.J., Bowman, K.P. et al. 1994. Arctic sea ice variability: Model sensitivities and a multidecadal simulation. *J. Geophys. Res.* **99** (C1): 919–936. http://dx.doi.org/10.1029/93JC02564.

Chaudhary, A. S. 2011. *Shale Oil Production Performance from a Stimulated Reservoir Volume*. MS Thesis, Texas A&M University, Texas (August 2011).

Chen, C., Balhoff, M. T., and Mohanty, K. K. 2014. Effect of Reservoir Heterogeneity on Primary Recovery and CO_2 Huff "n" Puff Recovery in Shale-Oil Reservoirs. *SPE Res Eval & Eng* **17** (3): 404-413. SPE-164553-PA. http://dx.doi.org/10.2118/164553-PA.

Chen, C. and Gu, M. 2017. Investigation of cyclic CO_2 huff-and-puff recovery in shale oil reservoirs using reservoir simulation and sensitivity analysis. *Fuel* **188**: 102-111. https://doi.org/10.1016/j.fuel.2016.10.006.

Chen, K. 2013. *Evaluation of EOR potential by gas and water flooding in shale oil reservoirs*. MS Thesis, Texas Tech University, Texas (2013).

Chen, R.B., Wang, W., and Wu, C.F.J. 2010. Building surrogates with overcomplete bases in computer experiments with applications to bistable laser diodes. *IIE Transactions* **43** (1): 39-53. http://dx.doi.org/10.1080/0740817X.2010.504686.

Chen, V.C.P., Tsui, K.L., Barton R.R. et al. 2006. A review on design, modeling and applications of computer experiments. *IIE Transactions*, **38** (4), 273–291. http://dx.doi.org/10.1080/07408170500232495.

Cheng, B. and Titterington, D.M. 1994. Neural networks: a review from a statistical perspective. *Statist. Sci*. **9** (1): 2–30. http://dx.doi.org/10.1214/ss/1177010638.

Cherian, B.V., Nichols, C.M., Panjaitan, M.L. et al. 2013. Asset Development Drivers in the Bakken and Three Forks. SPE-163855-MS. http://dx.doi.org/10.2118/163855-MS.

Chidi, A., Xie, J., Pivarnik, A. et al. 2014. Application of Design of Experiment Workflow to the Economic Evaluation of an Unconventional Resource Play. SPE-169834-MS. http://dx.doi.org/10.2118/169834-MS.

Clark, A. J. 2009. Determination of Recovery Factor in the Bakken Formation, Mountrail County, ND. SPE- 133719-STU. http://dx.doi.org/10.2118/133719-STU.

Clarke, S.M., Griebsch, J.H. and Simpson, T.W. 2005. Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses. *J. Mech. Des* **127** (6): 1077-1087. http://dx.doi.org/10.1115/1.1897403.

Cobianco, S., Bartosek, M., Lezzi, A. et al. 1999. How to Manage Drill-In Fluid Composition to Minimize Fluid Losses during Drilling Operations. SPE-57581-MS. http://dx.doi.org/10.2118/57581-MS.

Coleman, D.E. and Montgomery, D.C. 1993. A systematic approach to planning for a designed industrial experiment. *Technometrics* **35** (1): 1-27.

Czitrom, V. 1999. One-Factor-at-a-Time versus Designed Experiments. *The American Statistician* **53** (2): 126 - 131. http://dx.doi.org/10.2307/2685731.

Damsleth, E., Hage, A., and Volden, R. 1992. Maximum Information at Minimum Cost: A North Sea Field Development Study with an Experimental Design. *J Pet Technol* **44** (12): 1350-1356. SPE- 23139-PA. http://dx.doi.org/10.2118/23139-PA.

Daniel, C. 1973. One-at-a-Time Plans. *Journal of the American Statistical Association* **68** (342): 353-360. http://dx.doi.org/10.2307/2284076.

Devegowda, D. and Gao, H. 2007. Integrated Uncertainty Assessment for Unconventional Gas Reservoir Project Development. SPE-111203-MS. http://dx.doi.org/10.2118/111203-MS.

Dombrowski, R.J., Fonseca, E.R., Karanikas, J.M. et al. 2015. Fluid Injection in Light Tight Oil Reservoirs. US Patent 9127544.

Dong, C. and Hoffman, B.T. 2013. Modeling Gas Injection into Shale Oil Reservoirs in the Sanish Field, North Dakota. SPE- 168827-MS. http://dx.doi.org/10.1190/URTEC2013-185.

Dyn, N., Levin, D., and Rippa, S. 1986. Numerical procedures for surface fitting of scattered data by radial basis functions. *SIAM J. Sci. and Stat. Comput.* **7** (2): 639–659. http://dx.doi.org/10.1137/0907043.

Egeland, T., Holden, L., and Larsen, E.A. 1992. Designing Better Decisions. SPE-24275-MS. http://dx.doi.org/10.2118/24275-MS.

EIA. 2013. Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formations in 41 Countries outside the United States. Washington, D.C.

EIA. 2016a. Annual Energy Outlook. Washington, D.C.

EIA. 2016b. Short-Term Energy Outlook. Washington, D.C.

EIA. 2016c. Trends in U.S. Oil and Natural Gas Upstream Costs. Washington, D.C.

Eriksson, L. 2008. *Design of Experiments: Principles and Applications*, 3rd edition. Sweden: MKS Umetrics AB.

Eshkalak, M.O., Al-Shalabi, E.W., Sanaei, A. et al. 2014. Simulation study on the CO_2-driven enhanced gas recovery with sequestration versus the re-fracturing treatment of horizontal wells in the U.S. unconventional shale reservoirs. *Natural Gas Science and Engineering* **21**: 1015–1024. http://dx.doi.org/10.1016/j.jngse.2014.10.013.

Evans, B. and Fisher, D. 1994. Overcoming Process Delays with Decision Tree Induction. *IEEE Expert* **9** (1): 60-66. http://dx.doi.org/10.1109/64.295130.

Fai-Yengo, V., Rahnema, H., and Alfi, M. 2014. Impact of Light Component Stripping During CO_2 Injection in Bakken Formation. SPE-1922932-MS. http://dx.doi.org/10.15530/urtec-2014-1922932.

Fisher, R.A. and Bennett, J.H., and Yates, F. 1990. *Statistical Methods, Experimental Design and Scientific Inference*, 1st edition. NY: Oxford University Press.

Fragoso, A., Trick, M., Harding, T. et al. 2017. Coupling of Wellbore and Surface Facilities Models with Reservoir Simulation to Optimize Recovery of Liquids from Shale Reservoirs. SPE-185079-MS. http://dx.doi.org/10.2118/185079-MS.

Fragoso, A., Wang, Y., Jing, G. et al. 2015. Improving Recovery of Liquids from Shales through Gas Recycling and Dry Gas Injection. SPE-177278-MS. http://dx.doi.org/10.2118/177278-MS.

Frey, D.D., Engelhardt, F., and Greitzer, E.M. 2003. A role for "one-factor-at-a-time" experimentation in parameter design. *Research in Engineering Design* **14** (2): 65–74. http://dx.doi.org/10.1007/s00163-002-0026-9.

Frey, D.D. and Wang, H. 2006. Adaptive One-Factor-at-a-Time Experimentation and Expected Value of Improvement. *Technometrics* **48** (3): 418-431. http://dx.doi.org/10.1198/004017006000000075.

Friedman, J.H. 1991. Multivariate adaptive regression splines. *Annals Statistics* **19** (1): 1–141. http://dx.doi.org/10.1214/aos/1176347963.

Gamadi, T. 2014. *Experimental and Numerical Study of the EOR Potential in Shale Oil Reservoirs by Cyclic Gas Injection*. PhD dissertation, Texas Tech University, Texas (2014).

Gamadi, T.D., Elldakli, F., and Sheng J.J. 2014. Compositional Simulation Evaluation of EOR Potential in Shale Oil Reservoirs by Cyclic Natural Gas Injection. SPE-2014-1922690-MS. http://dx.doi.org/10.15530/urtec-2014-1922690.

Ghaderi, S.M., Clarkson, C.R., and Chen, Y. 2012. Optimization of WAG Process for Coupled CO_2 EOR-Storage in Tight Oil Formations: An Experimental Design Approach. SPE-161884-MS. http://dx.doi.org/10.2118/161884-MS.

Giunta, A. and Watson, L.T. 1998. A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models. *Proc*., 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization, St. Louis, MO, 2-4 September, AIAA-98-4758, Vol. 1, 392-404. http://dx.doi.org/10.2514/6.1998-4758.

Giunta, A., Wojtkiewicz, S. and Eldred, M. 2003. Overview of Modern Design of Experiments Methods for Computational Simulations. AIAA 2003-649. http://dx.doi.org/10.2514/6.2003-649.

Goodwin, N. 2015. Bridging the Gap Between Deterministic and Probabilistic Uncertainty Quantification Using Advanced Proxy Based Methods. SPE-173301-MS. http://dx.doi.org/10.2118/173301-MS.

Goodwin, N. and Powell, M. 2012. Simulation and Uncertainty: Lessons from Other Industries. *The Way Ahead* **8** (3): 22-24. SPE-0312-020-TWA. http://dx.doi.org/10.2118/0312-020-TWA.

Green, S. 1991. How many subjects does it take to do a regression analysis. *Multivariate Behav Res* **26** (3): 499-510. http://dx.doi.org/10.1207/s15327906mbr2603_7.

Gregory, G.A. 1974. Application of Factorial Design Analysis to Producing Well Pressure-Drop Modelling. *J Can Pet Technol* **13** (2): 21-27. PETSOC-74-02-01. http://dx.doi.org/10.2118/74-02-01.

Gupta, R., Van Elk, J.F., Tjia, D. et al. 2012. Experimental Design Methodology for Reserves Quantifications Based on Soft Computing Modelling Methods. SPE-160364-MS. http://dx.doi.org/10.2118/160364-MS.

Haight, J.M. 2010. Research-to-Practice in a Safety Context. ASSE-10-A5.

Hamlin, H.S., Smye, K., Dommisse, R., et al. 2017. Geology and Petrophysics of the Bakken Unconventional Petroleum System. URTEC-2670679-MS. http://dx.doi.org/10.15530/urtec-2017-2670679.

Han, L. and Gu, Y. 2014. Optimization of Miscible CO_2 Water-Alternating-Gas Injection in the Bakken Formation. *Energy Fuels* **28** (11): 6811–6819. http://dx.doi.org/10.1021/ef501547x.

Hardy, R.L. 1971. Multiquadratic equations of topography and other irregular surfaces. *J. Geophys. Res.* **76** (8): 1905–1915. http://dx.doi.org/10.1029/JB076i008p01905.

Harrell, F.E. Jr. 2001. *Regression modeling strategies*. NY: Springer.

Harrington, E.C. 1965. The Desirability Function. *Industrial Quality Control* **21** (10): 494-498.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Cary, NC: SAS Institute Inc.

Hibbert, D.B. 2012. Experimental design in chromatography: A tutorial review. *Journal of Chromatography B* **910**: 2-13. http://dx.doi.org/10.1016/j.jchromb.2012.01.020.

Hoffman, B. T. 2012. Comparison of Various Gases for Enhanced Recovery from Shale Oil Reservoirs. SPE-154329-MS. http://dx.doi.org/10.2118/154329-MS.

Iwere, F. O., Heim, R. N., and Cherian, B. V. 2012. Numerical Simulation of Enhanced Oil Recovery in the Middle Bakken and Upper Three Forks Tight Oil Reservoirs of the Williston Basin. SPE-154937-MS. http://dx.doi.org/10.2118/154937-MS.

Izadi, G., Zhong, M., and LaFollette, R.F. 2013. Application of Multivariate Analysis and Geographic Information Systems Pattern-Recognition Analysis to Production Results in the Bakken Light Tight Oil Play. SPE-163852-MS. http://dx.doi.org/10.2118/163852-MS.

Jain, S., Soliman, M., Bokane, A. et al. 2013. Proppant Distribution in Multistage Hydraulic Fractured Wells: A Large-Scale Inside-Casing Investigation. SPE-163856-MS. http://dx.doi.org/10.2118/163856-MS.

Jamshidnezhad, M. 2015. *Experimental Design in Petroleum Reservoir Studies*. Gulf Professional Publishing.

Jin, R., Chen, W. and Simpson, T. 2001. Comparative studies of metamodelling techniques under multiple modelling criteria. *Struct. Multidisc. Optim.* **23** (1): 1 – 13. http://dx.doi.org/10.1007/s00158-001-0160-4.

Jin, R., Chen, W., and Sudjianto, A. 2002. On Sequential Sampling for Global Metamodeling for in Engineering Design. *Proc*., ASME 2002 Design Engineering

Technical Conferences and Computer and Information in Engineering Conference, Montreal, Canada, September 29-October 2, DETC2002/DAC-34092, 539-548. http://dx.doi.org/10.1115/DETC2002/DAC-34092.

Johnson, R.T., Montgomery, D.C., and Jones, B. 2011. An empirical study of the prediction performance of space-filling designs. *International Journal Experimental Design and Process Optimization* **2** (1): 1-18. http://dx.doi.org/10.1504/IJEDPO.2011.038048.

Jones, B. and Goos, P. 2012. I-optimal versus D-optimal split-plot Response surface designs. *Journal of Quality Technology*, **44** (2), 85–101.

Jones, D.R., Schonlau, M., and Welch, W.J. 1998. Efficient global optimization of expensive black–box functions. *Journal of Global Optimization* **13** (4): 455–492. http://dx.doi.org/10.1023/A:1008306431147.

Joseph, V.R. 2016. Space-filling designs for computer experiments: A review. *Quality Engineering* **28** (1): 28-35. http://dx.doi.org/10.1080/08982112.2015.1100447.

Joshi, S. 2014. EOR: Next Frontier for Unconventional Oil. *J Pet Technol*, 20-22. SPE - 0614-0020-JPT. http://dx.doi.org/10.2118/0614-0020-JPT.

Joslin, K., Ghedan, S.G., Abraham, et al. 2017. EOR in Tight Reservoirs, Technical and Economical Feasibility. SPE-185037-MS. http://dx.doi.org/10.2118/185037-MS.

Kalla, S. and White, C.D. 2007. Efficient Design of Reservoir Simulation Studies for Development and Optimization. *SPE Res Eval & Eng* **10** (6): 629-637. SPE-95456-PA. http://dx.doi.org/10.2118/95456-PA.

Kalra, S. and Wu, X. 2015. Numerical Simulation Study on Injection for Enhancing Hydrocarbon Recovery and CO_2 Sequestration in Tight Oil Formations. American Rock Mechanics Association. ARMA 15-520.

Kanfar, M.S., Ghaderi, S.M., Clarkson, et al. 2017. A Modeling Study of EOR Potential for CO_2 Huff-n-Puff in Tight Oil Reservoirs - Example from the Bakken Formation. SPE-185026-MS. http://dx.doi.org/10.2118/185026-MS.

Kapustka, N., Boyd, C., Conrardy, C. et al. 2005. Parameter Characterization of Mechanically-Controlled Pulsed-Short Circuit Gmaw. ISOPE-I-05-368.

Kathel, P. and Mohanty, K.K. 2013. EOR in Tight Oil Reservoirs through Wettability Alteration. SPE-166281-MS. http://dx.doi.org/10.2118/166281-MS.

Kaufman, M., Balabanov, V., Giunta, A.A., et al. 1996. Variable complexity response surface approximations for wing structural weight in HSCT design. *Computational Mechanics* **18** (2): 112-126. http://dx.doi.org/10.1007/BF00350530.

Kelton, W.D. 1999. Designing Simulation Experiments. *Proc*., 1999 IEEE Winter Simulation Conference, 5-8 December, 33-38. http://dx.doi.org/10.1109/WSC.1999.823049.

Kemler, E.N. 1946. Factors Influencing the Leakage Resistance of Threaded Pipe Joints. API-46-275.

Kenett, R., Zacks, S., and Amberti, D. 2014. *Modern Industrial Statistics: with Applications in R, MINITAB and JMP*, 2nd edition. Chichester, UK: John Wiley & Sons.

Klasi, M.L. 1980. On the Optimum Design of Rock Mechanics Parameter Studies with Numerical Models. ARMA-80-0566.

Kleijnen, J.P.C. 1998. Experimental Design for Sensitivity Analysis Optimization and Validation of Simulation Models. In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, edition, Banks, J., Chap. 6, 173-223. John Wiley & Sons.

Kleijnen, J.P.C. 2008. *Design and analysis of simulation experiments*, 2nd edition. NY: Springer.

Kleijnen, J.P.C. 2010. Sensitivity Analysis of Simulation Models: an Overview. *Procedia - Social and Behavioral Sciences,* **2** (6), 7585-7586. http://dx.doi.org/10.1016/j.sbspro.2010.05.130.

Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., et al. 2005. State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments. *Journal on Computing* **17** (3): 263–289. http://dx.doi.org/10.1287/ijoc.1050.0136.

Klimberg, R. and McCullough, B.D. 2016. Fundamentals of Predictive Analytics with JMP®, 2nd edition. Cary, NC: SAS Institute Inc.

Kong, B., Wang, S., and Chen, S. 2016. Simulation and Optimization of CO2 Huff-and-Puff Processes in Tight Oil Reservoirs. SPE-179668-MS. http://dx.doi.org/10.2118/179668-MS.

Koselka, R. 1996. The new mantra: MVT. Forbes, March 11, 114 – 118.

Kothari, C.R. 2004. *Research Methodology Methods and Techniques*, 2nd revised edition. New Delhi: New Age International Publishers.

Krueger, R.F. 1956. Joint Bullet and Jet Perforation Tests (Progress Report). API-56-126.

Kuhn, P.P., di Primio, R. and Horsfield, B. 2010. Bulk composition and phase behavior of petroleum sourced by the Bakken Formation of the Williston Basin. *Proc*. of the 7th Petroleum Geology Conference, London, March 30 – April 2, 2009, p. 1065–1077.

Kurtoglu, B. 2014. *Integrated reservoir characterization and modeling in support of enhanced oil recovery for Bakken*. PhD dissertation, Colorado School of Mines, Colorado (2014).

Kurz, B.A., Schmidt, D.D., and Cortese, P.E. 2013. Investigation of Improved Conductivity and Proppant Applications in the Bakken Formation. SPE-163849-MS. http://dx.doi.org/10.2118/163849-MS.

Lafollette, R., Holcomb, W.D., and Aragon, J. 2012. Impact of Completion System, Staging, and Hydraulic Fracturing Trends in the Bakken Formation of the Eastern Williston Basin. SPE-152530-MS. http://dx.doi.org/10.2118/152530-MS.

Law, A.M. 2014. A Tutorial on Design of Experiments for Simulation Modeling. *Proc*., 2014 IEEE Winter Simulation Conference, 7-10 December, 66-80. http://dx.doi.org/10.1109/WSC.2014.7019878.

Lawal, K.A. 2009. Modelling Subsurface Uncertainties with Experimental Design: Some Arguments of Non-Conformists. SPE-128350-MS. http://dx.doi.org/10.2118/128350-MS.

Lechner, J.P., and Zangl, G. 2005. Treating Uncertainties in Reservoir Performance Prediction with Neural Networks. SPE-94357-MS. http://dx.doi.org/10.2118/94357-MS.

Leedy, P.D. and Ormrod, J.E. 2013. *Practical Research: Planning and Design*, 10th edition. Boston: Pearson.

Lewis, S.M. and Dean, A.M. 2001. Detection of Interactions in Experiments on Large Numbers of Factors. *J. R. Statist. Soc. B.* **63** (4): 633-672. http://dx.doi.org/10.1111/1467-9868.00304.

Li, B. and Friedmann, F. 2005a. Novel Multiple Resolutions Design of Experiment/Response Surface Methodology for Uncertainty Analysis of Reservoir Simulation Forecasts. SPE- 92853-MS. http://dx.doi.org/10.2118/92853-MS.

Li, B. and Friedmann, F. 2005b. A Novel Response Surface Methodology Based on "Amplitude Factor" Analysis for Modeling Nonlinear Responses Caused by Both Reservoir and Controllable Factors. SPE-95283-MS. http://dx.doi.org/10.2118/95283-MS.

Li, H., Sarma, P., and Zhang, D. 2011. A Comparative Study of the Probabilistic-Collocation and Experimental-Design Methods for Petroleum-Reservoir Uncertainty Quantification. *SPE J.* **16** (2): 429-439. SPE-140738-PA. http://dx.doi.org/10.2118/140738-PA.

Li, L., Sheng, J.J., and Sheng, J. 2016. Optimization of Huff-n-Puff Gas Injection to Enhance Oil Recovery in Shale Reservoirs. SPE-180219-MS. http://dx.doi.org/10.2118/180219-MS.

Li, L., Sheng, J.J., Watson, M., et al. 2015. Experimental and Numerical Upscale Study of Cyclic Methane Injection to Enhance Shale Oil Recovery. AIChE 425710.

Li, L., Sheng, J.J., and Xu, J. 2017. Gas Selection for Huff-n-Puff EOR in Shale Oil Reservoirs Based upon Experimental and Numerical Study.SPE-185066-MS. http://dx.doi.org/10.2118/185066-MS.

Lin, Y. 2004. *An Efficient Robust Concept Exploration Method and Sequential Exploratory Experimental Design*, PhD Dissertation, Georgia Institute of Technology, Atlanta (2004).

Lin,Y., Mistree, F., Allen, J.K. et al. 2004. Sequential metamodeling in engineering design. *Proc.*, 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY, August 30 – September 1, AIAA-2004-4304. http://dx.doi.org/10.2514/6.2004-4304.

Ling, K., Wu, X., Han, G. et al. 2016. Optimising the Multistage Fracturing Interval for Horizontal Wells in Bakken and Three Forks Formations. SPE-181788-MS. http://dx.doi.org/10.2118/181788-MS.

Liu, L. 2005. Could Enough Samples be More Important than Better Designs for Computer Experiments? *Proc*., 38th Annual Simulation Symposium, San Diego, CA, 4-6 April, 107-115. http://dx.doi.org/10.1109/ANSS.2005.17.

Liu, G., Sorensen, J.A., Braunberger, J.R. et al. 2014. CO_2-Based Enhanced Oil Recovery from Unconventional Reservoirs: A Case Study of the Bakken Formation. SPE-168979-MS. http://dx.doi.org/10.2118/168979-MS.

Loeppky, J.L., Sacks, J., and Welch, W.J. 2009. Choosing the sample size of a computer experiment: A practical guide. *Technometrics* **51** (4): 366–376. http://dx.doi.org/10.1198/TECH.2009.08040.

Lolon, E., Hamidieh, K., Weijers, L., et al. 2016. Evaluating the Relationship between Well Parameters and Production Using Multivariate Statistical Models: A Middle Bakken and Three Forks Case History. SPE-179171-MS. http://dx.doi.org/10.2118/179171-MS.

Ludvigsen, B.E. and Le, H. 2015. DST Matching and Interpretation through Numerical Well Testing on the Johan Sverdrup Field. SPE-175088-MS. http://dx.doi.org/10.2118/175088-MS.

Lye, L.M. 2002. Design of experiments in civil engineering: are we still in the 1920's? *Proc*., Annual Conference of the Canadian Society for Civil Engineering, Montreal, Canada, 5-8 June.

Mathews, P.G. 2005. *Design of Experiments with MINITAB*, 1st edition. Wisconsin: ASQ Quality Press.

Maugeri, L. 2012. Oil: The Next Revolution. Discussion Paper 2012-10. Belfer Center for Science and International Affairs, Harvard Kennedy School.

McKay, M.D., Beckman, R.J., and Conover, W.J. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21** (2): 239–245.

Mishra, S., Ganesh, P.R., Schuetter, J., et al. 2015. Developing and Validating Simplified Predictive Models for CO2 Geologic Sequestration. SPE-175097-MS. http://dx.doi.org/10.2118/175097-MS.

Mohaghegh, S.D., Gaskari, R., and Maysami, M. 2017. Shale Analytics: Making Production and Operational Decisions Based on Facts: A Case Study in Marcellus Shale. SPE-184822-MS. http://dx.doi.org/10.2118/184822-MS.

Montgomery, D.C. 2012. *Design and Analysis of Experiments*, 8th edition. Hoboken, NJ: John Wiley & Sons.

Morris, M.D. 2006. An Overview of Group Factor Screening. In *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, ed. Dean, A. and Lewis, S., Chap. 9, 191-206, NY: Springer.

Morsy, S. 2014. *Experimental and Simulation Study of Improved Oil Recovery in Shale Formations*. PhD dissertation, Texas Tech University, Texas (2014).

Morsy, S., Sheng, J.J., and Soliman, M.Y. 2013. Waterflooding in the Eagle Ford Shale Formation: Experimental and Simulation Study. SPE-167056-MS. http://dx.doi.org/10.2118/167056-MS.

Myers, R.H., Montgomery, D.C., and Anderson-Cook, C.M. 2009. *Response surface methodology: process and product optimization using designed experiments*, 3rd edition. Hoboken, NJ: John Wiley & Sons.

National Academy of Engineering 1995. *Forces Shaping the U.S. Academic Engineering Research Enterprise*. Washington D.C.: National Academy Press. https://doi.org/10.17226/4933.

Nguyen, N.T., Dang, C.T., Chen, Z. et al. 2015. Optimization of Hydraulic Fracturing Design with Future EOR Considerations in Shale Oil Reservoirs. SPE-174307-MS. http://dx.doi.org/10.2118/174307-MS.

Nojabaei, B., Johns, R.T., and Chu, L. 2012. Effect of Capillary Pressure on Fluid Density and Phase Behavior in Tight Rocks and Shales. SPE-159258-MS. http://dx.doi.org/10.2118/159258-MS.

NIST/SEMATECH. 2016. *E-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook. (accessed 15 October 2016).

Osterloh, W.T., Mims, D.S., and Meddaugh, W.S. 2013. Probabilistic Forecasting and Model Validation for the First-Eocene Large-Scale Pilot Steamflood, Partitioned Zone, Saudi Arabia and Kuwait. *SPE Res Eval & Eng* **16** (1): 97-116. SPE-150580-PA. http://dx.doi.org/10.2118/150580-PA.

Peng, C.Y. and Gupta, R. 2003. Experimental Design in Deterministic Modelling: Assessing Significant Uncertainties. SPE-80537-MS. http://dx.doi.org/10.2118/80537-MS.

Peng, C.Y. and Gupta, R. 2004. Experimental Design and Analysis Methods in Multiple Deterministic Modelling for Quantifying Hydrocarbon In-Place Probability Distribution Curve. SPE-87002-MS. http://dx.doi.org/10.2118/87002-MS.

Popov, E., Myasnkov, A., Cheremisin, A. et al. 2016. Experimental and Computational Complex for Determination of the Effectiveness of Cyclic Carbon Dioxide Injection for Tight Oil Reservoirs. SPE-181918-MS. http://dx.doi.org/10.2118/181918-MS.

Pu, W. 2013. *EOS Modeling and Reservoir Simulation Study of Bakken Gas Injection Improved Oil Recovery in the Elm Coulee Field, Montana*. PhD dissertation, Colorado School of Mines, Colorado (2013).

Pu, H., and Li, Y. 2015. CO_2 EOR Mechanisms in Bakken Shale Oil Reservoirs. Carbon Management Technology Conference. CMTC-439769-MS. http://dx.doi.org/10.7122/439769-MS.

Rivera, D. S., Mohanty, K., and Balhoff, M. 2015. Reservoir simulation and optimization of Huff-and-Puff operations in the Bakken Shale. *Fuel* **147**: 82-94. http://dx.doi.org/10.1016/j.fuel.2014.12.062.

Rosenberg, H.S. and Syrett, B.C. 1996. Effect of Environment on Coatings in FGD Ductwork. NACE-96451.

Sacks, J., Schiller, S.B. and Welch, W.J. 1989a. Designs for computer experiments. *Technometrics* **31** (1): 41–47. http://dx.doi.org/10.1080/00401706.1989.10488474.

Sacks J., Welch, W.J., Mitchell, T.J., et al. 1989b. Design and Analysis of Computer Experiments. *Statistical Science* **4** (4): 409-423. http://dx.doi.org/10.1214/ss/1177012413.

Sall, J., Creighton, L. and Lehman, A. 2007. *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP*, 4th edition. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2015. *JMP® 12 Design of Experiments Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2016. *JMP® 13 Predictive and Specialized Modeling*, 2nd edition. Cary, NC: SAS Institute Inc.

Sanchez, S.M. 2005. Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments. *Proc*., 2005 IEEE Winter Simulation Conference, 4-7 December, 69-82. http://dx.doi.org/10.1109/WSC.2005.1574241.

Santner, T.J., Williams, B.J., and Notz, W.I. 2003. *The Design and Analysis of Computer Experiments*, 1st edition. NY: Springer.

Saputelli, L., Lopez, C., Chacon, A., et al. 2014. Design Optimization of Horizontal Wells with Multiple Hydraulic Fractures in the Bakken Shale. SPE-167770-MS. http://dx.doi.org/10.2118/167770-MS.

Sarg, J.F. 2012. The Bakken – An Unconventional Petroleum and Reservoir System. Department of Energy (DoE).

Sasena, M., Parkinson, M., Goovaerts, P., et al. 2002. Adaptive Experimental Design Applied to An Ergonomics Testing Procedure. *Proc*., ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference, Montreal, Canada, September 29-October 2, DETC2002/DAC-34091, 529-537. http://dx.doi.org/10.1115/DETC2002/DAC-34091.

Sawyer, D.N., Cobb, W.M., Stalkup, F.I. et al. 1974. Factorial Design Analysis of Wet-Combustion Drive. *SPE J.* **14** (1): 25-34. SPE-4140-PA. http://dx.doi.org/10.2118/4140-PA.

Saxena, U., and Vjekoslav, P. 1971. Factorial Designs as an Effective Tool in Mining and Petroleum Engineering. SPE-3333-MS.

Schafer, J.L. and Graham, J.W. 2002. Missing data: our view of the state of the art. *Psychol Methods* **7**: 147-177.

Schlumberger. 2013. Taking Centre Stage: Pushing the Stage Count Limit in Bakken Wells. Oilfield Technology.

Schneider, J. and Moore, A. 2000. A Locally Weighted Learning Tutorial using Vizier 1.0. CMU-RI-TR-00-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (February 2000).

Schuetter, J., and Mishra, S. 2015. Experimental Design or Monte Carlo Simulation? Strategies for Building Robust Surrogate Models. SPE-174905-MS. http://dx.doi.org/10.2118/174905-MS.

Schulze-Riegert, R., Nwakile, M., Skripkin, S., et al. 2016. Scalability and Performance Efficiency of History Matching Workflows using MCMC and Adjoint Techniques Applied to the Norne North Sea Reservoir Case Study. SPE-180106-MS. http://dx.doi.org/10.2118/180106-MS.

Schmidt, F.L. 1971. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educ Psychol Meas* **31** (3): 699-714. http://dx.doi.org/10.1177/001316447103100310.

Shmueli, G., Bruce, P.C., Stephens, M.L., et al. 2017. *Data Mining for Business Analytics: Concepts, Techniques, and Applications with JMP Pro*. Hoboken, NJ: John Wiley & Sons.

Schweitzer, R. and Bilgesu, H.I. 2009. The Role of Economics on Well and Fracture Design Completions of Marcellus Shale Wells. SPE-125975-MS. http://dx.doi.org/10.2118/125975-MS.

Shams, M. 2016. Reservoir Simulation Assisted History Matching: From Theory to Design. SPE-182808-MS. http://dx.doi.org/10.2118/182808-MS.

Sheather, S. 2009. *A Modern Approach to Regression with R*. Verlag NY: Springer.

Sheng, J.J. 2015. Enhanced oil recovery in shale reservoirs by gas injection. *Natural Gas Science and Engineering* **22**: 252-259. http://dx.doi.org/10.1016/j.jngse.2014.12.002.

Sheng, J.J., and Chen, K. 2014. Evaluation of the EOR potential of gas and water injection in shale oil reservoirs. *Unconventional Oil and Gas Resources* **5**: 1-9. http://dx.doi.org/10.1016/j.juogr.2013.12.001.

Sheng, J.J., Cook, T., Barnes, W. et al. 2015. Screening of the EOR Potential of a Wolfcamp Shale Oil Reservoir. ARMA-2015-438.

Shoaib, S. and Hoffman, B.T. 2009. CO_2 Flooding the Elm Coulee Field. SPE-123176-MS. http://dx.doi.org/10.2118/123176-MS.

Simpson, T.W. 1998. Comparison of response surface and kriging models in the multidisciplinary design of an aerospike nozzle. ICASE Report No. 98-16, NASA (February 1998).

Simpson, T.W., Lin, D.K.J., and Chen, W. 2001a. Sampling strategies for computer experiments: design and analysis. *International Journal of Reliability and Applications* **2** (3): 209–240.

Simpson, T.W., Mauery, T.M., and Korte, J.J. et al. 2001b. Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA Journal* **39** (12): 2233-2241. http://dx.doi.org/10.2514/2.1234.

Simpson, T.W., Peplinski, J.D., Koch, P.N. et al. 1997. On the Use of Statistics in Design and the Implications for Deterministic Computer Experiments. *Proc*., 1997 ASME Design Engineering Technical Conference, Sacramento, CA, 14-17 September.

Simpson, T.W., Peplinski, J.D., Koch, P.N. et al. 2001c. Metamodels for Computer-based Engineering Design: Survey and recommendations. *Engineering with Computer* **17** (2): 129–150. http://dx.doi.org/10.1007/PL00007198.

Smith, M. 1993. *Neural networks for statistical modeling*, 3rd edition. NY: Von Nostrand Reinhold.

Taber, J.J. 1983. Technical Screening Guides for the Enhanced Recovery of Oil. SPE-12069. http://dx.doi.org/10.2118/12069-MS.

Taber, J.J., Martin, F.D., and Seright, R.S. 1997. EOR Screening Criteria Revisited - Part 1: Introduction to Screening Criteria and Enhanced Recovery Field Projects. *SPE Reservoir Engineering* **12** (3): 189-198. http://dx.doi.org/10.2118/35385-PA.

Tanco, M., Viles, E., Álvarez, M.J. et al. 2010. Why is not design of experiments widely used by engineers in Europe? *Journal of Applied Statistics 3***7** (12): 1961–1977. http://dx.doi.org/10.1080/02664760903207308.

Trocine, L. and Malone, L.C. 2001. An overview of newer, advanced screening methods for the initial phase in an experimental design. *Proc*., 2001 IEEE Winter Simulation Conference, 9-12 December, 749-754. http://dx.doi.org/10.1109/WSC.2001.977263.

Varadarajan, S., Chen, W. and Pelka, C. 2000. The Robust Concept Exploration Method with Enhanced Model Approximation Capabilities. *Engineering Optimization* **32** (3): 309-334. http://dx.doi.org/10.1080/03052150008941302.

Vinassa, M., Cudjoe, S.E., Gomes, J.H.B., et al. 2015. A Comprehensive Approach to Sweet-Spot Mapping for Hydraulic Fracturing and CO_2 Huff-n-Puff Injection in Chattanooga Shale Formation. SPE-175952-MS. http://dx.doi.org/10.2118/175952-MS.

Vogel, L.C. 1956. A Method for Analyzing Multiple Factor Experiments - Its Application to a Study of Gun Perforating Methods. SPE-727-G. http://dx.doi.org/10.2118/727-G.

Wan, H., Ankenman, B.E., and Nelson B.L. 2006. Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation. *Operations Research* **54** (4): 743-755. http://dx.doi.org/10.1287/opre.1060.0311.

Wan, T., Meng, X., Sheng, J.J. et al. 2014. Compositional Modeling of EOR Process in Stimulated Shale Oil Reservoirs by Cyclic Gas Injection. SPE-169069-MS. http://dx.doi.org/10.2118/169069-MS.

Wan, T., Sheng J.J., and Soliman, M.Y. 2013a. Evaluation of the EOR Potential in Shale Oil Reservoirs by Cyclic Gas Injection. SPWLA-2013-MM.

Wan, T., Sheng J.J., and Soliman, M.Y. 2013b. Evaluate EOR Potential in Fractured Shale Oil Reservoirs by Cyclic Gas Injection. SPE-168880-MS. http://dx.doi.org/10.1190/URTEC2013-187.

Wan, T., Yu, Y., Sheng, J. J. 2015. Experimental and numerical study of the EOR potential in liquid-rich shales by cyclic gas injection. *Unconventional Oil and Gas Resources* **12**: 56-67.

Wang, G.G. 2003. Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points. *J. Mech. Des* **125** (2): 210-220. http://dx.doi.org/10.1115/1.1561044.

Wang, S., and Chen, S. 2016a. Evaluation and Prediction of Hydraulic Fractured Well Performance in Montney Formations Using a Data-Driven Approach. SPE-180416-MS. http://dx.doi.org/10.2118/180416-MS.

Wang, S., and Chen, S. 2016b. A Comprehensive Evaluation of Well Completion and Production Performance in Bakken Shale Using Data-Driven Approaches. SPE-181803-MS. http://dx.doi.org/10.2118/181803-MS.

Wang, G., Dong, Z., and Aitchison, P. 2001. Adaptive Response Surface Method - A Global Optimization Scheme for Computation-intensive Design Problems. *Journal of Engineering Optimization* **33** (6): 707-734.

Wang, G.G. and Shan, S. 2006. Review of Metamodeling Techniques in Support of Engineering Design Optimization. *J. Mech. Des* **129** (4) 370-380. http://dx.doi.org/10.1115/1.2429697.

Wang, F. and White, C.D. 2002. Designed Simulation for a Detailed 3D Turbidite Reservoir Model. SPE-75515-MS. http://dx.doi.org/10.2118/75515-MS.

Wang, X., Luo, P., Er, V. et al. 2010. Assessment of CO_2 Flooding Potential for Bakken Formation, Saskatchewan. SPE-137728-MS. http://dx.doi.org/10.2118/137728-MS.

Weissman, S.A. and Anderson, N.G. 2015. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev*. **19** (11): 1605–1633. http://dx.doi.org/10.1021/op500169m.

Wescott, A. 2016. *Reservoir characterization of the Middle Bakken member, Fort Berthold region, North Dakota, Williston Basin*, PhD Dissertation, Colorado School of Mines, Colorado (2004).

White, C.D. and Royer, S.A. 2003. Experimental Design as a Framework for Reservoir Studies. SPE-79676-MS. http://dx.doi.org/10.2118/79676-MS.

Winestock, A.G., and Colpitts, G.P. 1965. Advances in Estimating Gas Well Deliverability. *J Can Pet Technol* **4** (3): 111-119. PETSOC-65-03-01. http://dx.doi.org/10.2118/65-03-01.

Wolff, M. 2010. Probabilistic Subsurface Forecasting. SPE-132957-MS.

Wu, J. and Hamada, M. 2009. *Experiments: Planning, Analysis, and Optimization*, 2nd edition. Hoboken, NJ: John Wiley & Sons.

Xu, T., and Hoffman, T. 2013. Hydraulic Fracture Orientation for Miscible Gas Injection EOR in Unconventional Oil Reservoirs. SPE-168774-MS. http://dx.doi.org/URTEC2013-189.

Yamada, S. 2008. A Grammar of Design of Experiments in Computer Simulation. In *The Grammar of Technology Development*, ed. Tsubaki, H., Yamada, S., and Nishina, K., Chap. 7, 85-103, Japan: Springer.

Ye, X. 2016. *Optimum Hydraulic Fracture Conductivity for Shale/Tight Reservoirs*. MS thesis, University of Calgary, Calgary (June 2016).

Yeten, B., Castellini, A., Guyaguler, B. et al. 2005. A Comparison Study on Experimental Design and Response Surface Methodologies. SPE-93347-MS. http://dx.doi.org/10.2118/93347-MS.

Yu, W., Lashgari, H., and Sepehrnoori, K. 2014. Simulation Study of CO_2 Huff-n-Puff Process in Bakken Tight Oil Reservoirs. SPE-169575-MS. http://dx.doi.org/10.2118/169575-MS.

Yu, W. and Sepehrnoori, K. 2014a. An Efficient Reservoir Simulation Approach to Design and Optimize Unconventional Gas Production. Journal of Canadian Petroleum Technology 53 (2), 109 - 121. SPE-165343-PA. http://dx.doi.org/10.2118/165343-PA.

Yu, W. and Sepehrnoori, K. 2014b. Optimization of Well Spacing for Bakken Tight Oil Reservoirs. SPE- 2014-1922108-MS. http://dx.doi.org/10.15530/urtec-2014-1922108.

Zheng, J., Leung, J.Y., Sawatzky, R. P., et al. 2016. A Proxy Model for Predicting SAGD Production from Reservoirs Containing Shale Barriers. SPE-180715-MS. http://dx.doi.org/10.2118/180715-MS.

Zhu, P., Balhoff, M.T., and Mohanty, K.K. 2015. Simulation of Fracture-to-Fracture Gas Injection in an Oil-Rich Shale. SPE-175131-MS. http://dx.doi.org/10.2118/175131-MS.

Zhu, P., Balhoff, M.T., and Mohanty, K.K. 2017. Compositional modeling of fracture-to-fracture miscible gas injection in an oil-rich shale. *Petroleum Science and Engineering* **152**: 628–638. https://doi.org/10.1016/j.petrol.2017.01.031.

Zubarev, D.I. 2009. Pros and Cons of Applying Proxy-models as a Substitute for Full Reservoir Simulations. SPE-124815-MS. http://dx.doi.org/10.2118/124815-MS.

APPENDIX A

SUMMARY OF SOME DOE WORK DONE IN PETROLEUM ENGINEERING LITERATURE AND OTHER DISCIPLINES

Reference	Proxy	DoE	Notes
Peng and Gupta 2003	QPR	FuF, MC, & various versions of PB & FrF	• Two-level FrF. (e.g., PB) is enough for screening • Recommends adding control & axial points to two-level FrF • Recommends using expert knowledge to get efficient DoE (min. number of runs) for best proxy
Peng and Gupta 2004	QPR & KG	FuF, FrF, & LHD	• LHD did not perform better than three-level FrF • KG is not better than QPR
Li and Firedmann 2005b	Amplitude factor RSM, QPR, & TPS	CCD	• Amplitude factor RSM is more efficient & accurate than TPS & QPR for highly nonlinear problems without the need to partition the variable space
Yeten et al. 2005	QPR, KG, & TPS	SFD, PB, CCD, & D-opt	• Better results with SFDs than classical designs • QPR gives effect estimates that agree with other proxies
Kalla and White 2007	QPR & KG	OA, NOA, HS, & LHD	• OA & NOA are better than HS & LHD for sampling for uncertainty analysis • HS is more efficient than LHD for sampling for uncertainty analysis, especially in high dimensional problems • QPR is better than KG
Devegowda and Gao 2007	QPR (for CCD) & KG (for LHD)	CCD & LHD	• LHD/KG provides more accurate estimates than the traditional CCD/QPR through the whole variable space
Zubarev 2009	QPR, KG, TPS, & ANN	LHD	• Metamodeling method itself is not very influential on proxy performance as long as adequate input data is available • Using proxies in high dimensions is not recommended • KG & TPS are better for highly nonlinear response with KG being better for more nonlinearity
Li et al. 2011	QPR, LPR, LIPR, ANN, TPS, & KG	D-opt, exhaustive sampling, folded PB, & SFD	• Probabilistic collocation method (PCM) is more efficient & more accurate than tested DoEs for uncertainty analysis since most DoEs do not honor the probability distribution of factors • Various DoEs & proxies give different results • The best DoE & proxy are problem-specific

Table A.1—Summary of DoE and proxy techniques in petroleum engineering studies.[3]

[3] DoE abbreviations: AP = augmented pairs, BB = Box Behnken, CCD = central composite design, D-opt = D-optimal, FrF = fractional factorial, FuF = full factorial, HS = Hammersley sequence, LHD = Latin hyper-cube design, MC = Monte Carlo sampling, ME = maximum entropy, MM = MaxiMin, NOA = nearly orthogonal array, OA = orthogonal array, PB = Plackett-Burman, and SFD = space-filling designs.
Proxy abbreviations: ANN = artificial neural networks, KG = kriging, LIPR = polynomial regression (PR) with linear and interaction terms, LPR = linear PR, QPR = quadratic PR, and TPS = thin plate splines.

Reference	Proxy	DoE	Notes
Gupta et al. 2012	QPR, KG, ANN, & regression based decision tree	case 1 (2 factors): LHD & FuF case 2 (8 factors): LHD & D-opt	By comparing RMSE they concluded that: • For small no. of factors, FuF/D-opt & QPR are preferred • For large no. of factors, LHD & KG preferred • KG, ANN, & regression decision tree work best with LHD
Lechner, and Zangl 2005	ANN	Draper & Lin, BB, FrF, & various versions of CCD	• Recommends three levels not two
Mishra et al. 2015	QPR (for BB) & KG (for LHD)	BB & LHD (MM)	• LHD with KG gives better results than BB with QPR
Schuetter and Mishra 2015	QPR (for BB & AP) & KG (for LHD & ME)	BB, LHD (MM), ME, & AP	• LHD with KG & QPR have the best performance • LHD & ME with KG give better results than classical DoEs with regression • LHD with QPR is recommended since KG is more computationally costly (despite being slightly better) than QPR

Table A.1 Continued—Summary of DoE and proxy techniques in petroleum engineering studies.

Reference	Proxy	DoE	Notes
Simpson 1998 and Simpson et al. 2001b	KG & QPR	OA	• KG has same performance as second order QPR • Despite being complex in fitting, using, & validating, KG has the potential of higher accuracy, which justifies more research
Jin et al. 2001	QPR, MARS, RBF, & KG	LHD with different sample sizes	• RBF is usually the most accurate and robust method. PR is very accurate for nonsevere nonlinearity • QPR is the most efficient in terms of computational cost & KG is the least • QPR & MARS have good transparency (providing information about contribution of model terms), while KG & RBF do not • QPR & RBF are the easiest to use • Sampling technique affects the performance of metamodels
Simpson et al. 1997, 2001c (review papers)			• Building PR is manageable for up to 20 factors. However, obtaining data is the limiting factor for more than 10 factors. PR is simple & well-established but breaks down in case of high non-linearity • ANN can handle highly nonlinear & high dimensionality problems (10,000 parameters) if resources are available • Inductive learning suites discrete-valued factors & responses • KG is very flexible but complex & computationally expensive. It can either honor the data or smooth the data points & is suitable for deterministic modeling with up to 50 factors
Giunta and Watson 1998	KG and QPR	D-opt for 5- & 10-factor problems	• QPR is more accurate than KG even for highly non-quadratic cases

Table A.2—Summary of DoE and proxy techniques in other disciplines.[4]

[4] DoE abbreviations: CCD = central composite design, D-opt = D-optimal, GP IMSE = Gaussian process integrated mean square error, HS = Hammersley sequence, LHD = Latin hyper-cube design, MC = Monte Carlo sampling, OA = orthogonal array, RND = random sampling, SP = sphere packing, and UD = uniform design.
Proxy abbreviations: ANN = artificial neural networks, KG = kriging, LIP = least interpolating polynomials, MARS = multivariate adaptive regression splines, PR = polynomial regression, QPR = quadratic PR, RBF = radial basis function, RSM = response surface method, SVR = support vector regression, and TPS = thin plate splines.

Reference	Proxy	DoE	Notes
Simpson et al. 2001a	QPR, MARS, RBF, and KG	LHD, HS, OA, RND, & UD for 3- & 14-factor problems	• HS & UD give more accurate models with HS performing better with larger sample size while UD performs well with small sizes • KG & RBF give better accuracy for different DoEs & sample sizes • MARS is the least robust since its performance varies widely with different DoEs & sample sizes • PR gives average results & are good with low-order non-linearity • In general, accuracy increases by larger sizes. This is not true, however, when non-linearity is not severe (except for MARS)
Johnson et al. 2011	KG	LHD, UD, GP IMSE, & SP	• No design has an advantage over another regarding prediction accuracy • Adequate sample size of space-filling designs is the best way to enhance the predictive accuracy
Chen et al. 2006	QPR, ANN, MARS, KG, RBF, & LIP	UD, LHD, OA, OA-based LHD, HS, & FCD for 2- & 10-factor problems	• No DoE or metamodel outperforms the others. However, OA design & KG & RBF metamodels have consistent good performance • RSM designs perform well only for RSM models. Other designs are generally good for all metamodels
Sacks et al. 1989a	KG	CCD, LHD, & optimal	• CCD & LHD perform well with large sample size while optimal designs are good with small sample size
Clarke et al. 2005	KG, RBF, MARS, QPR, & SVR	CCD, inscribed CCD, HS, LHD, OA-based LHD, & UD for 2-, 3-, & 4-factor problems	• SVR is more accurate & more robust than the other used techniques • SVR combines the good predictive accuracy & robustness of KG with transparency & computational efficiency close to those of PR & RBF • SVR has a great potential as a metamodeling technique

Table A.2 Continued—Summary of DoE and proxy techniques in other disciplines.